Author:
Paterson, Alex R

Title:
The Hunt for Novel Prognostic Blood Biomarkers of Hypertension

The Hunt for Novel Prognostic Blood Biomarkers of Hypertension

University of Bristol

Alex R. Paterson

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Translational Health Sciences

September 2019

Abstract

Hypertension is defined as a chronic state of elevated blood pressure, and results in a significantly increased risk of developing diseases such as; coronary heart disease, kidney damage and stroke. Essential hypertension is a multifactorial and polygenic condition that remains asymptomatic until much of the damage has already been done, resulting in many going undiagnosed (1 in 3) until the health impact has become severe. Globally, hypertension is responsible for 9.4 million deaths every year. Therefore, in order to improve upon the diagnostic assessment of the disease, this current study made use of the Spontaneously Hypertensive Rat (SHR) as a polygenic model of elevated blood pressure. Using an optimised next generation sequencing approach to guide biomarker discovery, transcripts predicted to be significantly regulated between the SHR and its normotensive control were selected as candidate biomarkers of hypertension. The interferon-induced protein with tetratricopeptide repeats 1 (Ifit1) saw a robust elevation in the blood of SHRs at both juvenile and adult ages. This, coupled with its key integretive role in innate immunity, lead to Ifit1 being selected for further validation through a series of environmentally induced forms of elevating blood pressure. Ifit1 performed well as a biomarker of the SHR strain, against a host of normotensive controls. As such, Ifit1 was selected for additional characterisation in the blood through Flow Cytometry as well as assessment in a key integrative region of the brain tasked with cardiovascular control.

For Mum and Dad,

I could not have done it without your genetic material

Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Signed:.....Date:.....

Acknowledgements

I would like to take this opportunity to thank Prof. Charles Hindmarch and Prof. David Murphy for their continued mentorship. Together, they have inspired and motivated me throughout this body of work, and have left an indelible mark on my future career in research.

A sincere thank you also goes to Dr. Olivera Šarenac and her colleagues in Belgrade, for warmly welcoming me several times into their laboratory. Olivera's patience and knowledge have given me a firm foundation of surgical techniques in assessing physiology.

I would like to dedicate this work to my family, for their encouragement and support. They have always been a great source of strength for me and I will always admire them for this.

A special thank you goes to Dr Michael Greenwood and Dr Mingkwan Greenwood for their advice over the years. Not only have they supported my work at the experimental level, but they have always guided my research and provided a daily sounding board for ideas.

Despite their best efforts to distract and therefore derail my career, I would like to thank Ben Gillard, Dr Ben Flynn, George Horn, Dr Matt Birnie, Audrys Pauza, and Dr Paul Bishop. Their humour and ability to lift my spirits have seen me through the many stressful moments of the past four years.

Finally, I would like to extend a big thank you to the Murphy Group as a whole, in addition to many members of the Lightman Group. From experimental advice to coffee runs, the group have been there for any and all PhD woes. All of whom I am proud to call friends.

Contents

List of Figures

List of Tables

Acronyms

1K1C one-kidney one-clip.

2K1C two-kidney one-clip.

ACI August & Copenhagen inbred.

ANOVA analysis of variance.

ANS autonomic nervous system.

ASPA Home Office animals (scientific procedures) act 1986.

BALF broncho-alveolar lavage fluid.

BBDP biobreeding diabetes prone.

BHR borderline hypertensive rat.

BN Brown norway.

BP blood pressure.

BPM beats per minute.

BRS baroreceptor sensitivity.

BSA bovine serum albumin.

CI confidence interval.

CNS central nervous system.

CV coefficient of variance.

DBP diastolic blood pressure.

DESeq differential gene expression analysis based on the negative binomial distribution.

DGE differential gene expression.

DH dehydration.

dNTP deoxynucleotide triphosphate.

DP dorsal parvocellular subdivision.

EdgeR empirical analysis of digital gene expression data in R.

EDTA ethylenediaminetetraacetic acid.

ELISA Enzyme-Linked Immunosorbent Assay.

eQTL expressed quantitative trait-loci.

EtOH ethanol.

F344 Fischer 344.

FACS fluorescence-activated cell sorting.

FDR false discovery rate.

FFT fast fourier transformation.

FHH Fawn hooded high blood pressure.

FHL Fawn hooded low blood pressure.

FPKM fragments per kilobase of exon model per million reads.

GC guanine-cytosine.

GK Goto-Kakizaki.

GTF gene transfer format.

GWAS genome wide Association study.

HF high frequency.

Ifit1 interferon-induced protein with tetratricopeptide repeats 1.

ISG interferon stimulated gene.

ISRE interferon stimulated response elements.

L-NAME N(γ)-nitro-L-arginine methyl ester.

LE Long Evans.

LEW Lewis.

LF low frequency.

LH Lister hooded.

LH Lyon hypertensive.

LL Lyon low blood pressure.

LN Lyon normotensive.

LPS lipopolysaccharide.

MAP mean arterial pressure.

MCC Matthew's correlation coefficient.

MDS multidimensional scaling analysis.

MHS Milan hypertensive strain.

MIQE minimum information for publication of quantitative real-time PCR experiments.

mmHg millimetres of mercury.

MMP maximum mappable prefix.

MN magnocellular division.

MNS Milan normotensive strain.

MP medial parvocellular subdivision.

mRNA messenger ribonucleic acid.

NIL neurintermediate lobe.

NOS nitric oxide synthase.

NTS nucleus tractus solitarii.

P periventricular parvocellular subdividion.

PAMP pathogen associated molecular pattern.

PBS phosphate buffered saline.

PCA principal component analysis.

PCR polymerase chain reaction.

PI pulse interval.

PRRs pathogen-recognition receptor families.

PVN paraventricular nucleus.

QC quality control.

QUOD quality in organ donation.

RAAS renin-angiotensin-aldosterone-system.

RMA robust multiarray averaging.

RNA ribonucleic acid.

RPKM reads per kilobase of exon model per million reads.

rRNA ribosomal ribonucleic acid.

RVLM rostral ventrolateral medulla.

SBH Sabra hypertension prone.

SBN Sabra normotensive.

SBP systolic blood pressure.

SBS sequencing by synthesis.

SD Sprague-Dawley.

SEM standard error of the mean.

SHR spontaneously hypertensive rat.

SNP single nucleotide polymorphism.

SNS sympathetic nervous system.

SNV single nucleotide variants.

SON supraoptic nucleus.

SR salt resistant.

SS salt sensitive.

STAR spliced transcripts alignment to a reference.

STDev standard deviation.

TAC transcriptome analysis console.

TLR4 toll-like receptor 4.

TLRs toll-Like Receptors.

TMM trimmed mean of the M-values.

TPM transcripts per million.

VLF very low frequency.

VP ventral parvocellular subdivision.

WAG Wistar albino Glaxo.

WHO world health organisation.

WKY Wistar Kyoto.

Chapter 1

Introduction

1.1 Hypertension

1.1.1 A Rise in Blood Pressure

In 1628, the English physician William Harvey published a pioneering description of the functionality of the heart and blood. In his seminal *Exercitatio Anatomica de Motu Cordis et Sanguinis in Animalibus* (Anatomical Account of the Motion of the Heart and Blood in Animals), Harvey was the first to observe the properties of the blood, and its integrated systemic circulation around the body by the heart. In doing so, he gave rise to the modern history of cardiovascular research and mankind's understanding of its importance (Figure 1.1). Prior to his work, blood was thought to be separated between venous and arterial systems that were quite separate, and the heart simply considered an internal regulator of heat. Over 100 years later, Stephen Hales took Harvey's work and further propelled the field with his first quantitative measurements of blood pressure.[2] Using finely produced glass columns inserted into the arteries of animal subjects, Hales could note the height to which the column of blood rose; giving rise to a comparative metric of blood pressure.

Historically, the disease of hypertension was described as a "fullness disease", resulting from an excess of blood in a sufferer's vessels. Early Persian medical texts

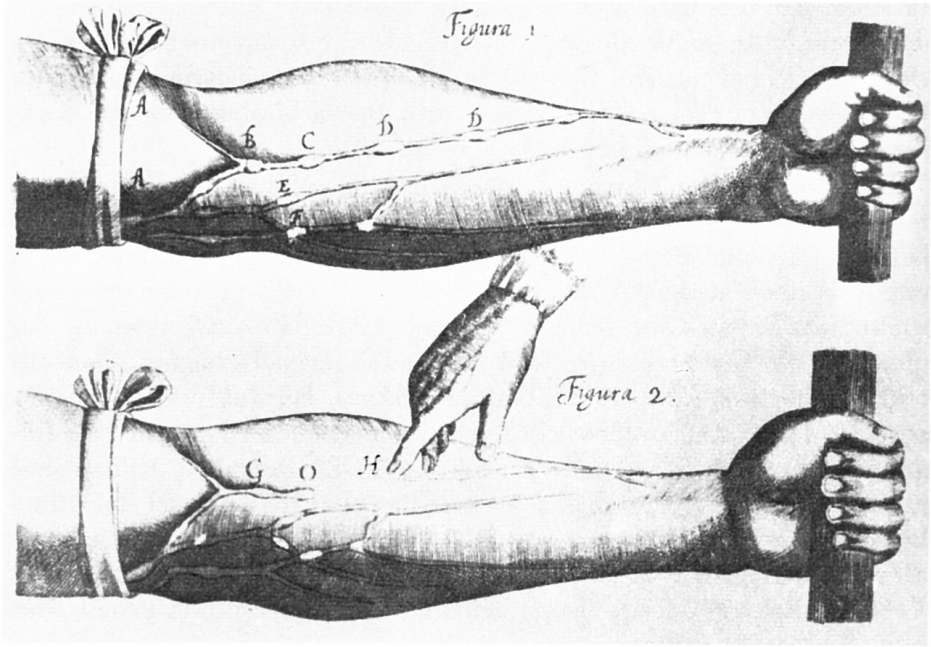


Figure 1.1: A drawing taken from Harvey's *Exercitatio Anatomica de Motu Cordis et Sanguinis in Animalibus* depicting an experiment on valves in the venous circulatory system.

described the symptoms as including; headaches, general redness, being warm to the touch, distension of the skin, vascular rupture, and even hemorrhagic stroke.[3] A natural treatment option therefore took the form of blood-letting or the application of leeches to balance the perceived overabundance of blood present.[4]

It was not until the late 19^{th}/ early 20^{th} century that hypertension gained true clinical importance, catalysed by the invention of the cuff-based sphygmomanometer by Scipione Riva-Rocci in 1896.[5] This novel device was expanded upon shortly after by Russian surgeon Nikolai Korotkov. Korotkov improved upon Riva-Rocci's approach through the use of a stethoscope to listen to a series of audible sounds when the arteries were auscultated alongside the deflating arm cuff of the sphygmomanometer. Korotkov noted these sounds as being indicative of changes in blood pressure when the heart muscles were contracted (Systole) and relaxed (Diastole), allowing for physicians to discern between the two in a clinical setting.[6] Armed with the knowledge of these new diagnostic "Korotkov sounds", as they become known, physicians could begin to pick apart the complex relationship between blood pressure and health; and in 1913,

Theodore Janeway became the first person to describe how individuals with an increased blood pressure were statistically more likely to die earlier than those with a lower blood pressure.[7]

Today, hypertension is clinically classed with the single diagnostic criteria of chronically elevated blood pressure (BP). However, in reality hypertension presents a multifactorial disease state with a significant impact on a range of physiological processes. BP is a normally distributed quantitative trait across the general population, with normal arterial BP (normotensive) falling within the range of 100 – 140mmHg at systole and 60 – 90mmHg at diastole. The disease of hypertension becomes diagnosed when blood pressure is measured above the upper end of the normal distribution over multiple time points. †

Current European guidelines define hypertension when measurements are consistently over 140/90mmHg systolic and diastolic respectively, whereas the latest American guidelines have moved this threshold down to over 130/80mmHg.[8,9] This new definition of hypertension within the American guidelines is based largely on the results derived from large meta-analyses such as SPRINT (Systolic Blood Pressure Intervention Trial).[10] In contrast to this, the european guidelines are based on population attributable risk statistics. However, both guidelines suggest a therapeutic end goal of a BP under 130/80mmHg. In 2003, an additional designation of "Pre-hypertension" was termed for individuals with a systolic blood pressure (SBP) between 120-139mmHg and a diastolic blood pressure (DBP) of between 80-89mmHg in order to catch those at risk of further elevations of BP.[11] This additional designation allows for prophylactic measures to be put in place in order to slow or stop the rise in BP before it leads to a higher risk of cardiovascular disease; either through clinical, or lifestyle intervention. In reality, these classifications provide an arbitrary series of thresholds against a normally distributed

†These multiple measurements are essential in ensuring elevated BP is a chronic symptom and not one resultant of a patient's anxiety at a clinical visit; a phenomenon known as "White coat syndrome".

range of BP within the population. As such, they only serve as a guide that may well be subject to change, as their primary purpose is to identify those at risk of the disease states associated with hypertension. Thereby enabling healthcare professionals to intervene accordingly.

As BP begins to rise, so too do the risk factors associated. Elevated blood pressure significantly increases one's susceptibility towards; stroke, ischemic heart disease, and peripheral vascular disease.[12–14] Additional associated diseases include; aortic aneurysms, diffuse aetherosclerosis, atrial fibrillation, pulmonary embolism, and chronic kidney disease.[15,16] These diseases are not limited to the heart and kidneys, as hypertension is also a serious risk factor for retinopathy, and even generalised cognitive impairment and dementia.[17,18] It is therefore not surprising that hypertension is classed as the most important preventable risk factor for premature death in the world.[19]

1.1.2 The Hypertensive Pandemic - A Global Burden

Globally, hypertension is responsible for approximately 9.4 million deaths every year and causes around 45% of heart disease related deaths and 51% of deaths due to stroke.[20] In 2008, around 40% of adults over 25 years old were diagnosed with hypertension, a significant rise from 600 million in 1980 to approximately 1 billion. This increasing prevalence has been attributed to an ever-aging population in addition to increased behavioural risk factors, such as poor diet or sedentary lifestyle.[21] The world health organisation (WHO) estimates that this number will continue rising to 1.5 billion by 2020.[22] As of 2001, sub-optimal blood pressure was estimated to cost healthcare providers $370 billion per year globally; a number that represents about 10% of the world's overall healthcare expenditure.[23]

1.1.3 A Genetic Basis for Hypertension

For a long time it was unknown whether the genetic basis for BP was a quantitative trait, or a dichotomous disease state. These perspectives gave rise to the famous Platt-Pickering debate during the 1950's, a discussion on the unimodal or bimodal distributions of BP in the general population. Platt regarded hypertension as a distinct condition, showing a bimodal distribution within the population.[24] Whereas Pickering saw BP as a continuous distribution, with some reaching the clinically set threshold levels of hypertension (Figure 1.2[25]). Platt determined that hypertension was a manifestation of simple Mendelian disease inheritance, grounded in a single genetic mutation that gives rise to the hypertensive phenotype. As the debate continued well into the 1960's, Pickering's quantitative model gained little support. It was not until increasing epidemiological evidence suggested the link between elevated BP and cardiovascular disease that this changed. The ability of a range of pharmaceutically distinct anti-hypertensive drugs to reduce BP, and therefore cardiovascular disease, all strengthened Pickering's normally distributed model. Furthermore, genome wide Association study (GWAS) cemented this within the scientific community, as they showed the importance of considering the polygenic association of genes with hypertension[26–28] †.

Today hypertension diagnoses are split into those where a discernible cause is known (Inessential or Secondary Hypertension), and those whereby the specific cause of hypertension can not be identified (Essential or Primary Hypertension). Secondary hypertension only makes up a small proportion of hypertensive cases, at between 5-10%.[11,31,32] It can be caused by a wide variety of conditions, including; renal disease, endocrine disorders, and side effects from pharmaceuticals. Due to the tangible nature of the disease, therapeutic measures can be put in place to target the cause of the disorder and effectively manage BP back to physiological levels.

The vast majority of hypertensive cases however (90-95%), have no known cause to

†It should be noted, that Platt's model cannot be entirely disregarded as there exist rare familial forms of hypertension with a Mendelian root, and therefore inherited as single-gene disorders.[29]

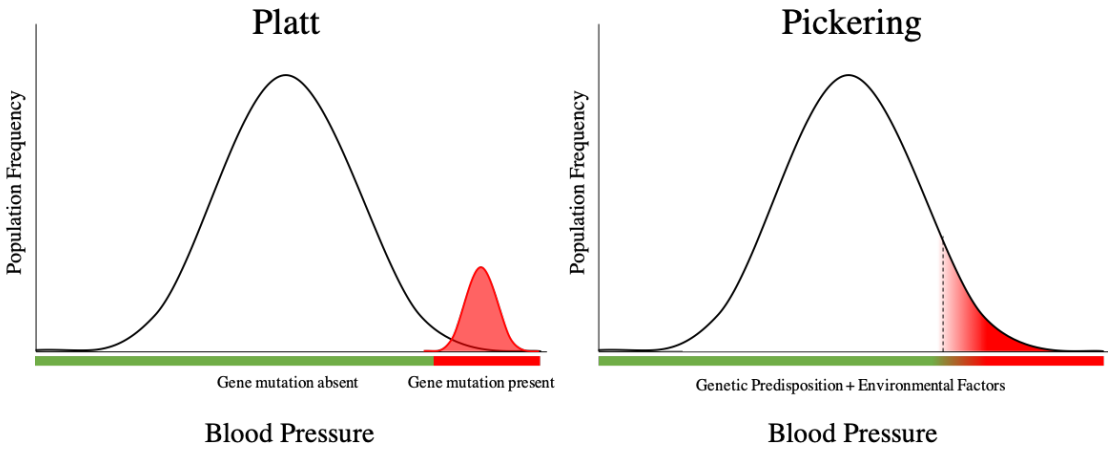


Figure 1.2: The Platt-Pickering debate over the qualitative or quantitative nature of hypertension. Platt considered hypertension to be a distinct condition, and therefore one resulting of Mendelian genetics. Whereas, Pickering saw blood pressure (BP) as a continuum amongst the general population, with hypertension arising in those at the extreme end of the distribution. He therefore concluded there to be a polygenic aetiology to the disease. Image adapted from Padmanabhan *et al.* 2012[30]

their chronically elevated BP. This grouping of individuals are clinically known as those with Primary Hypertension. Primary hypertension accounts for the highest proportion of cardiovascular disease and is considered to be polygenic in nature.[33]

BP regulation is a complex physiological process, governed by an intricate network of interacting pathways that culminate in changes to cardiac output or peripheral resistance. For these reasons, dissecting the genetic underpinnings of the hypertensive phenotype presents a unique set of challenges. Previous studies have demonstrated the genetic contribution towards BP management. Based on familial studies, specifically those focusing on monozygotic and dizygotic twins, it is estimated to have a heritability rate of between 31-68%.[34,35] It is for these cumulative reasons that isolating the precise genetic cause towards hypertension has proven so difficult.

1.1.4 The Brain's Role in Hypertension

The mechanisms by which humans regulate mean arterial BP are not completely understood. While they can be broadly grouped into distinct regulatory systems, in

reality they are interrelated. One key regulatory system is that of the renin-angiotensin-aldosterone-system (RAAS) which centres on production of angiotensin II, a potent endogenous vasoconstrictor.[36] Over the last few decades significant progress has been made to develop a series of treatments for hypertension, many focusing on modulation of the RAAS. However, around 40% of hypertensive patients remain unresponsive to pharmaceuticals of these classes.[37] This coupled with the established associations between an elevated sympathetic nervous system (SNS) and hypertension implicate the neurogenic basis of the disease. One key system at play is that of the baroreceptor reflex. Briefly, the baroreceptor reflex involves afferent projections from arterial baroreceptors to central nervous system (CNS) pathways (Figure 1.3). These in turn, effect their inhibitory or excitatory processes on sympathetic outflow, ultimately culminating in alterations to heart rate and cardiac output, in addition to influencing the long term activity of the RAAS.[38] The baroreceptor reflex is primarily tasked with short term BP fluctuations, responding to tissue's metabolic demand throughout a day. However, it is thought that chronic hypertension arises when alterations to the baseline of this reflex are modified, effectively resetting the homeostatic set-point of BP to an elevated mean arterial pressure.

For these reasons, the search for blood-born biomarkers has been supplemented by looking toward an additional CNS region of the hypothalamus involved in this pathway; the paraventricular nucleus (PVN). The PVN is located lateral to the third ventricle and can be further differentiated into eight distinct groups of nuclei in rats consisting of either large magnocellular or smaller parvocellular neurons.[40] These parvocellular projections directly influence sympathetic tone via the rostalventrolateral medulla (RVLM) and the intermediolateral cell column.[41] The magnocellular neurons are tasked with the production of oxytocin and vasopressin, the latter being key for osmoregulation via secretion from the pituitary. Moreover, the PVN is also involved in several other physiological functions, including; response to biological stressors and energy homeostasis.[41–43] All of which underpin these regions as key regions of interest

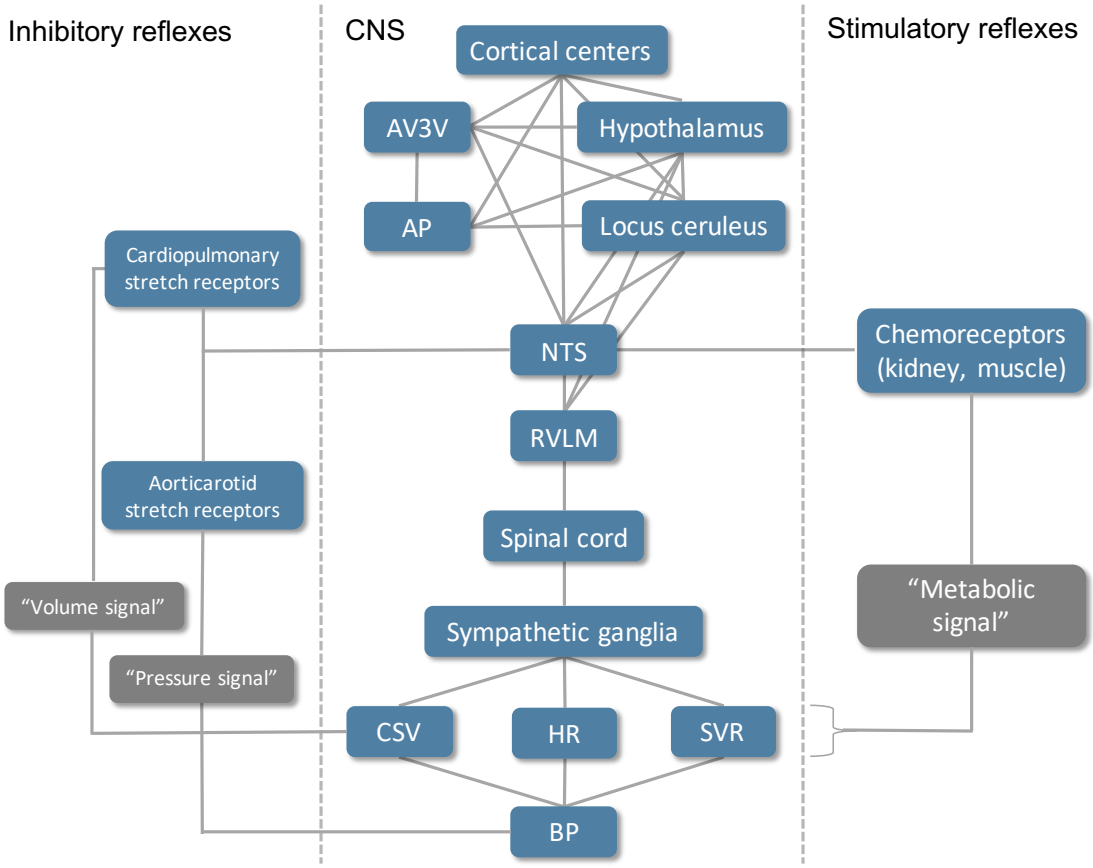


Figure 1.3: Schematic overview of the central nervous system (CNS) control of sympathetic outflow and ultimate regulation of BP. central nervous system (CNS) outflow is the result of integrated actions from a collection of CNS regions, including several centres of the cortex, lower centers of the hypothalamus, circumventricular organs (Area Postrema, AP; Periventricular Anteroventral Third Ventricle, AV3V), and basal ganglia (especially the Locus ceruleus). The critical integrator site for these regions is the nucleus tractussolitaries (NTS), situated within the medulla oblongata. This site receives inhibitory signals from baroreflexes (pressure and volume signals) and stimulatory signals from muscular and renal chemoreceptors (metabolic signals). The ultimate determiner of sympathetic nervous system (SNS) outflow is the rostal ventrolateral medulla (RVLM), which receives tonic inhibition from the adjacent NTS. The circumventricular organs, such as the AP and AV3V, are devoid of the blood brain barrier and therefore are of particular interest for their role in sensing hormonal and metabolic signals from the blood. The complex integrated signalling taking place centrally ultimately results in the RVLM sending signals via the spinal cord and sympathetic ganglia to regulate stroke volume (SV), heart rate (HR), and systemic vascular resistance (SVR). Image adapted from Chopra *et al.*, 2011[39]

when studying the neurogenic basis of hypertension.

1.2 Animal Models in Hypertension

When trying to elucidate the aetiology, and any biomarkers that should precede the onset of hypertension, animal models can prove invaluable. There exist several animal models of hypertension, each with their own aetiology and resulting disease characteristics. Therefore, a great level of care is required when selecting one to study that best reflects essential hypertension in humans.

1.2.1 Surgical Models of Hypertension

One commonly used model in hypertension research is that of the two-kidney one-clip (2K1C). The model requires surgically constricting one renal artery in the rat, leading to a chronically reduced renal perfusion with the contralateral kidney remaining unmodified. The model is characterised by an early and rapid rise in plasma renin in response to the reduced renal artery pressure followed by a resulting increase in circulating angiotensin II (ang II).[44] The 2K1C/1K1C models are considered robust models of hypertension. However both require invasive surgery to induce a hypertensive state putting a high degree of stress on the animal, and arguably not mimicking the aetiology of human hypertension.

1.2.2 Genetic Models of Hypertension

One of the most widely used rat models of hypertension is that of the spontaneously hypertensive rat (SHR). Originally developed in the 1960's by Okakamoto and Aoki, the strain was obtained by inbreeding Wistar rats that exhibited the highest BP. After only a few generations the result was a strain of rat that spontaneously begins to develop increased BP at around 5-6 weeks of age until plateauing at a systolic BP of

180-200mmHg. This is in contrast to the accepted normal blood pressure of between 80-100mmHg.[45] The importance of this strain resides in its onset of hypertension without the need for pharmacological, surgical or dietary intervention (contrary to models such as the Two-Kidney One-Clip or Dahl Salt-Sensitive rats).

For this reason, the SHR presents a model very distinct from the others; one that mimics the polygenic aspect of human hypertension. Furthermore, as with any accurate model of a disease state, fully hypertensive adult SHRs develop many of the hallmarks of end-organ damage associated with hypertension in humans. These include; cardiac failure (approx. 60% of SHRs develop heart failure after 18 months of age), cardiac hypertrophy (approx. 30% increase in size), and proteinuria from renal dysfunction.[46–48] The SHR also exhibits impaired endothelium dependent relaxation, a feature of human hypertension that is commonly exhibited and portends the cardiovascular risk that often occurs when left untreated.[49] Further support for using the SHR as a proxy to human hypertension is given by its responsiveness of many of the pharmaceutical classes that are used on humans. For example; renin-angiotensin-aldosterone-system (RAAS) inhibitors, calcium antagonists, and vasodilators.[50–52] However, SHRs are less responsive to diuretics and endothelin antagonists, and the data on the effect of beta-blockers has proven unequivocal.[50,52,53] The spontaneous nature of SHR hypertension coupled with its polygenic aetiology make it an ideal candidate to study the genetic basis of human hypertension. At 4 weeks old (juvenile), SHRs present a prehypertensive age, and therefore an age of interest when trying to assess key differences in their transcriptome that could predict their inevitable rise in BP.

1.2.3 Reconciliation of the Genome-Environment Interface

Unlike monogenic diseases, complex polygenic disease states are much more likely to be a function of the environment. Hypertension is no different in this respect, as multiple modifiable risk factors towards the disease have been identified. It is therefore important

to consider the interactions between a genetically predisposed individual, and the lifestyle factors that may contribute to a hypertensive phenotype. One such modifiable risk factor in the development of hypertension is salt intake.[54] While the specific guidelines towards recommended salt intake have varied in light of contradicting studies, there is a clear relationship between a high salt diet and elevated blood pressure.[55]

With this is mind the Dahl salt-sensitive rats represent another commonly used model of hypertension. In the 1950's, Meneely *et al.* began to study the effects of a high salt diet on the BP of rats. The team noted a linear relationship between the percentage of NaCl (between 0 and 10%) within the diet and the average BP measured within a cohort of Sprague-Dawley (SD). They further noted the marked degree of variation between individuals in their BP response to salt.[56] Then in the 1960's, Dr. Lewis Dahl built upon the observations of Meneely *et al.* by studying the effects of chronic excess of salt in rat models.[57] Dahl took advantage of the observation that rats had a differential response to salt. Being mindful of the data suggesting genetic influences on blood pressure in humans, he set about selectively breeding rats either susceptible (Dahl-Salt Sensitive) or resistant (Dahl-Salt Resistant) towards the hypertensive effects of a high salt diet (8% NaCl). Dahl *et al.* managed to clearly separate these lines after only three generations of breeding. The blood pressure of Dahl-Salt Resistant animals were effectively similar between those fed a control diet or a high salt diet. However the Dahl-Salt Sensitive animals develop severe and fatal hypertension under a high salt diet.[58,59]

Although the organ damage endpoints are comparable to many other models of hypertension, with cardiac hypertrophy (approx. 32%) and eventual cardiac failure (4-5 months of age) robustly observed, the model requires a dietary intervention before these become apparent.[46] The Dahl Salt-Sensitive and resistant animals constitute a hugely important model of studying blood pressure regulation, as they present an opportunity for reconciling the genome-environment interface in the aetiology of hypertension.

1.3 The Hunt for Biomarkers

Often referred to as a "silent killer", hypertension is asymptomatic in its early stages. This results in many patients remaining undiagnosed (1 in 3) until the impact becomes severe and possibly fatal, highlighting the need for improved diagnostics.[60] Should elevated BP remain unmanaged, it becomes a serious risk factor for the diseases outlined prior. In order to effectively manage hypertension, the need to intervene before associated risk factors develop is paramount. For that reason, it is important to identify a series of biological markers, or biomarkers, that precede the onset of the disease. These characteristic changes can be at any level of biology, provided they serve as an objective metric for the progression of the disease accurately and reproducibly. Another crucial consideration of a clinically relevant biomarker is its accessibility. Blood offers an accessible tissue that can be obtained through relatively non-invasive means. Here the search for prognostic biomarkers is centred on the biological level of the blood transcriptome. This is due to the ability of RNA to reflect the biological state of the individual in a highly dynamic manner; being able to integrate both the genetic and epigenetic mechanisms at play in the aetiology of the disease. Indeed many studies are looking towards transcript expression in offering a robust biomarker for a disease state due to their early and more accurate predictive properties.[61–63]

1.4 Transcriptomics

The transcriptome is the total sum of transcripts and their respective quantities within a cell. These can include all of the mRNAs, small RNAs and non-coding RNAs a cell is actively expressing. Focusing the search on the transcriptome allows for a dynamic metric of gene expression, allowing a reconciliation between an organism's genetic predisposition towards a disease, and how that is being affected due to environmental stimuli.

1.4.1 Evolution of Nucleotide Sequencing

In 2003, the complete mapping of the human genome gave rise to a new era in genetic analysis. The project sought to sequence a representative version of each human chromosome (approximately 3 billion bases in total).

The base resolution of libraries was resolved via a process based on Sanger Sequencing. From its conception in 1977, Frederick Sanger's eponymous sequencing technique was a ground breaking development in the ability to unravel the genetic code of life.[64] The central development in the technique was the use of specific chain-terminating inhibitors of deoxynucleotide Triphosphates (dNTPs) to prevent the elongation of oligonucleotides. DNA templates that were to be sequenced were combined with DNA polymerase, free deoxynucleotide bases (dATP, dCTP, dGTP, and dTTP), and separated out into 4 separate incubations, each with sparse mixture of one radioisotope labeled dideoxynucleotide base (ddATP, ddCTP, ddGTP, and ddTTP). Following a heating step to denature the template DNA, a cooling step allows the DNA primer to anneal. DNA polymerase could then synthesise a complementary DNA strand one base at a time. These bases are randomly incorporated until a dideoxynucleotide base was utilised, terminating the DNA elongation. The resulting mixture was one of newly synthesised DNA strands, each differing in length by a single nucleotide and all labeled at their 3' end with a radioisotope. In order to read the sequences researchers could separate the new DNA strands based on size through electrophoresis as each strand differed by a single DNA molecule. By running all strands through an electrophoresis gel, the resulting pattern of labeling on radiation sensitive film could be read one base at a time, taking note of which original incubation it belonged to to therefore to resolve the base for that read.[65]

The whole process was painstakingly slow, and to this day remains the world's largest collaborative biological project taking a total of 13 years to reach completion. This huge undertaking relied on a massively parallel and admirably driven workforce

consisting of 20 institutions that spanned 6 nations.[66] As a result, the whole process cost an estimated $2.7 billion to yield the first draft sequence.[65]

1.4.2 Next Generation Sequencing

Today an individual's genome can be resolved in as little as a day. This huge reduction in time is due to a coevolution of massively parallel sequencing technology, and an exponential development in the computational power required to process it. Furthermore, these rapid advances have resulted in a marked reduction in associated costs. The downward trend in the price to sequence the human genome has far surpassed Moore's law, a metric used to assess technological improvements as computational power advances. These advancements have been termed as "Next-Generation Sequencing (NGS)". They expand upon the procedural concepts outlined by Sanger Sequencing and translate them to a massively-parallel approach. While many variations of NGS technology exist, one of the main approaches employs the sequencing by synthesis (SBS) method.

The procedure begins with the random fragmentation of genomic DNA, or cDNA, to produce a pool (Library) of uniformly sized fragments. A series of adapters are ligated to each end of the fragments, as well as unique stretches of oligonucleotides for later identification of the original sample from which the fragments were derived ("Barcoding"). Each library is mixed and washed over a flow cell where a lawn of bound oligonucleotides bind the complementary adapters that flank the sample fragment. At this point cluster generation is required to yield clusters of similar fragments such that their cumulative signal is sufficient to be detected by the sequencer. This can take place within the sequencer machine (as an onboard cluster module), or on an external machine (such as the cBot, Illumina, Inc. - San Diego, CA, USA). This is achieved via a process known as "Bridge Amplification", and begins with the opposite end of a ligated fragment bending over and "bridging" to another complementary oligo bound

to the flow cell lawn. Through repeated denaturation and extension cycles (similar to a polymerase chain reaction (PCR)), fragments are locally amplified from single molecules into millions of unique, clonal clusters along the flow cell. At this point the SBS process can take place.

SBS technology classically uses four fluorescently labeled nucleotides to sequence millions of clusters bound to the flow cell in parallel [†]. During each of the sequencing cycles, a single complementary deoxynucleotide triphosphate (dNTP) binds to the oligonucleotide chain, and reversibly terminates strand elongation. After dNTP incorporation the bound fluorophore is excited and detected through laser imaging to resolve the specific base that was incorporated within that cluster. The reversible terminator dNTP can then be enzymatically cleaved ready for the next cycle of incorporation.

By the very nature of NGS, huge amounts of data are produced from any given experiment. Depending on the organism being studied, genomes can vary in size from the smallest non-viral genome at 160Kb, to the largest known at 670GB. The complexity of the transcriptome represents its own challenge with regions of the genome being expressed more than once; giving rise to an additional dimension of data to interpret.

The ability to deal with this scale of "Big Data" in an acceptable length of time is incredibly computationally demanding. To maintain a feasible time frame for any given project, focus needs to be given to optimising mathematical approaches to best optimise available computational power.

1.5 Bioinformatics Challenges

The chaining of processes required to take the raw .Fastq sequencing files to a final biologically interpretable output is referred to as a "Pipeline" (Figure 1.4). Currently,

[†]Illumina's NextSeq Series introduced the concept of 2-channel SBS technology, with only two flourophores required. Here only two images are needed to resolve each base, contrary to the classical 4 images employed before. This is achieved by using a red and green wavelength filter to identify colours as red (C), green (T), a combination (yellow, A) or no colour, given (G). This concept has radically sped up sequencing runs but has raised concerns about accuracy.

there are no "gold standards" in optimal pipelines for differential expression. Different laboratories favour different pipelines depending on the organism being studied, their research goals, or simply their familiarity with the software.

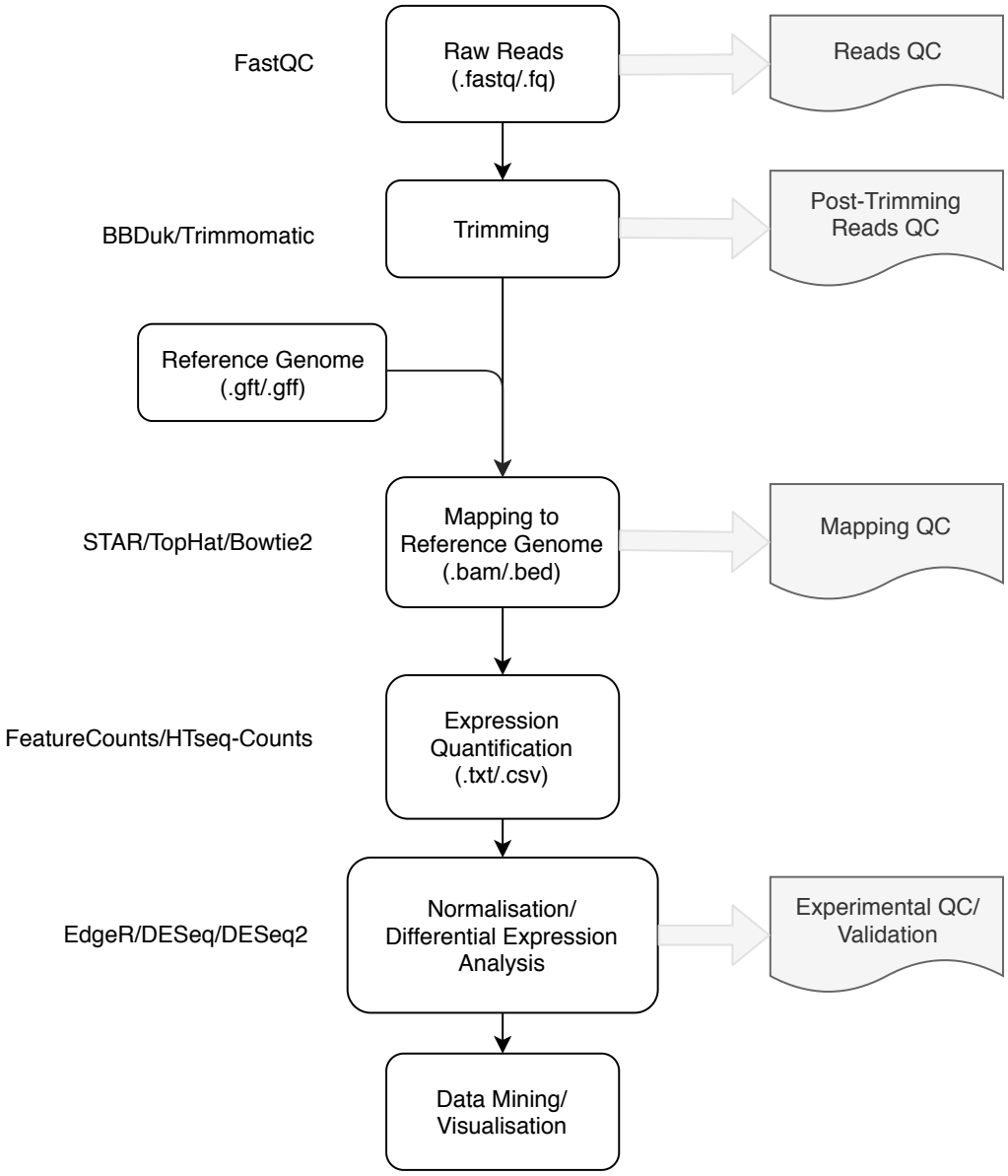


Figure 1.4: Example Pipeline for RNAseq Experiment. Here each step is shown from raw demultiplexed reads to final biologically interpretable output files. Some examples of software programs are shown for each step. Also displayed are the common file types associated with each stage in the pipeline.

1.5.1 Quality Control

Bioinformatic quality control (QC) of outputted NGS reads is equally as essential as the QC steps taken at the library preparation stage. These QC checks should be employed pertinently at several stages along the bioinformatic analysis. The very first step in any RNAseq pipeline is the QC analysis of the raw reads to assess; sequence quality, guanine-cytosine (GC) content, overrepresentation of particular subsequences (k-mers, where k equals the length of the sequence in bases), duplicated reads and the presence of the adapters used during the library preparation stage. Each of these metrics can be indicative of a potential error that may have confounding effects on the final results. To this end, FastQC provides a comprehensive software package for Illumina-generated reads to produce graphical plots of the above analyses.[67]

While the technology used for the Human Genome Project has since become outdated for experiments requiring a high-throughput approach, its legacy has given rise to a series of novel assessments that remain in use today. The most common metric for base calling accuracy is the Phred Quality score. The Phred score was originally developed as an algorithmic approach towards resolving Sanger sequencing reads, taking into account the peak resolution and shape of the detected fluorophore signal. While these attributes no longer exist in next-generation sequencing technologies, the process of generating Phred scores has remained largely the same. Using different sets of parameters, relevant to the specific sequencing chemistry being used, accuracies are correlated against a large lookup data set of known accuracies (Figure 1.5).

The resulting Q-Score is one that represents the probability of an incorrect base call after any given number of bases resolved (Table 1.1). Lower Q-scores can increase levels of false positive variant calling, resulting in inaccurate conclusions or higher costs due to additional validation experiments. Today, the "gold standard" of sequencing experiments is a Q-score of 30 (Q30), which translates to a probability of 1 potential

$$Q = \log_{10} p$$

Where; Q = Phred Quality Score
p = Probability of an incorrect base

Figure 1.5: Equation for calculating Phred Score. Here probabilities (p) are retrieved from pre-calculated lookup tables of known accuracies.

miscall in 1,000 bases.

Phred scores have proven to be incredibly consistent throughout a series of different sequencing technologies, leading to it becoming the industry standard for all commercial sequencing technologies.[68]

Phred Quality Score	**Probability of incorrect base call**	**Accuracy of base call**
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1,000	99.9%
40	1 in 10,000	99.99%

Table 1.1: Table of example Phred scores and their accompanying base calling accuracies.

1.5.2 Trimming

It is essential to remove the adapters added to each fragment during the library preparation stage. A trimming software package (e.g. Trimmomatic or BBDuk) can be used to trim low quality bases, trim adapter sequences, or simply discard poor quality reads. All of which will improve mappability to a reference genome. However, it must be noted that the more aggressive the trimming approach, the greater loss of potentially important information contained within the reads. For this reason, a balance needs to be struck in order to get the most from the sequencing experiment.

Furthermore due to a phenomenon known as "Phasing", read quality classically falls towards the 3' end of reads.[69] This occurs when the nucleotide blocker, employed during

each sequencing by synthesis chemistry cycle, fails to be removed after signal detection. During the next cycle of chemistry, a nucleotide is prevented from binding the template DNA fragment, and the old nucleotide is once again detected. The fluorescent signal of the old nucleotide most probably differs from the signal produced around it from homologous template strands and is therefore not a synchronous signal. For the rest of the sequencing procedure the result will be one cycle behind the detection signal of the other strands of that homologous cluster, polluting the signal detected by the sequencer's camera and preventing a high confidence base call for that particular cluster. This phenomenon occurs with a relatively low probability. However over time this can accumulate and begin to pollute the detected light signal more and more. This is especially true with increasing read lengths. Sequence quality plots can quickly identify if this quality becomes too low across the read allowing the bioinformatician to make the decision whether to remove these low-quality bases (Figure 1.6).

1.5.3 Read Alignment

Once reads have been filtered or trimmed for optimum quality, they can be aligned to either a reference genome or transcriptome. Either present a series of pro's and con's to be assessed on an experiment by experiment basis. An important metric for read alignment success is simply the percentage of reads successfully mapped to the reference genome/transcriptome. Depending on how stringent the aligner algorithm, this can provide an overall indicator of the sequencing accuracy, and whether or not any contaminating DNA is present. As an example, an expected rate of between 70 and 90% of RNAseq reads should map to the human genome with some of those reads mapping to several identical regions within the reference (known as "Multi-Mapping Reads").[70] It would be expected that a lower level of total reads map when mapping to the transcriptome, as unannoted regions will be lost. As a consequence more reads become defined as multi-mapping due to the number of reads falling onto shared exons

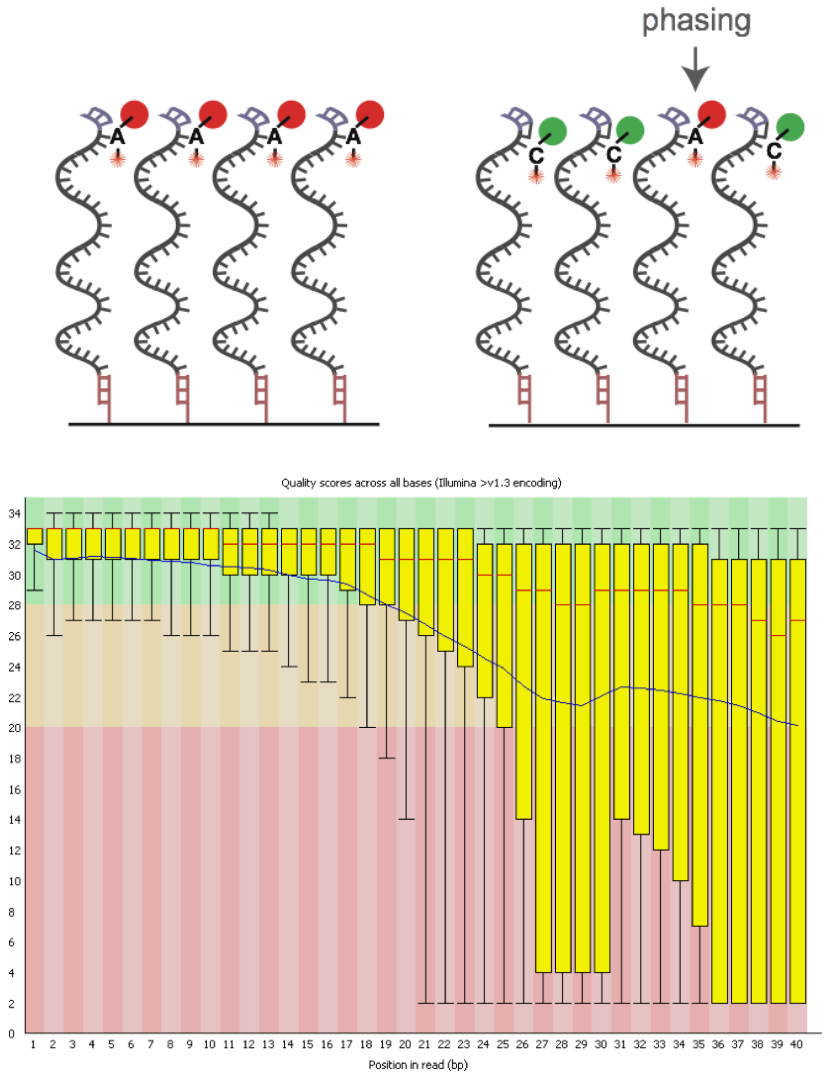


Figure 1.6: Phasing Phenomenon observed within Illumina's Sequencing by Synthesis technology (Illumina, Inc. - San Diego, CA, USA). Phasing occurs when a blocker nucleotide is not removed during one of the chemistry cycles of the experiment (Above, Image taken from ECSeq Bioinformatics[69]). Downstream signal is therefore asynchronous within the homologous cluster, and leads to a decreasing confidence in accurate base calling, as represented by a reduction in Phred score (Below, BoxWhisker type plot generated by FASTQC (v0.11.7)[67]).

from the same genes. For these reasons, priority has been given to mapping reads to the reference genome rather than the transcriptome.

Additional parameters that need to be considered are the uniformity of reads across exons, as reads that primarily accumulate at the 3' end of poly(A) selected samples may be indicative of low quality starting RNA.

1.5.4 Transcript Quantification and Normalisation

Most analyses of RNAseq data require an estimation of the number of genes expressed, usually for later differential gene expression (DGE) analysis. The most basic approach to quantifying expression levels is to simply aggregate the number of raw counts of mapped reads, through programs such as FeatureCounts, or HTSeq-count.[71,72] This process makes use of a gene transfer format (GTF) file, containing genomic coordinates for genes and their exons, to quantify the number of reads that fall within these boundaries. However, these "raw" read counts are not sufficient for comparing the expression of transcripts between samples, as they are greatly affected by a series of confounding factors inherent in both library preparation and RNAsequencing. These include factors such as; sequencing depth, transcript length, and inherent biases in RNAseq technlogy. Therefore, read normalisation is essential before read counts can be used downstream with any confidence.

The process of read normalisation is one that has presented the greatest variety in approaches. In order to convert raw read counts into a measure of relative abundance that can be biologically interpreted, normalisation is an essential step. It is this crucial nature that has led normalisation to gain so much research interest. As so many factors can lead to a potentially confounding effect on the final output, there is no single normalisation strategy that appears to adequately deal with all of these factors.

One of the early methods to address these confounders was to perform a within-sample normalisation, by dividing the raw count by the number of kilobases in its exon

model, and again by the total number of reads generated the sequencing experiment. The resulting measure was called the reads per kilobase of exon model per million reads (RPKM).[73] The RPKM, and its derivatives such as the analogous fragments per kilobase of exon model per million reads (FPKM), were for a long time the most frequently reported gene expression values for RNAseq [†]. The need to correct for gene length is only necessary for correctly assessing gene expression levels within the sample as compared with other genes. However, it it not necessary when comparing any changes to transcript expression within the same gene across samples as they will be subject to the same bias of longer genes accumulating more reads. For this reason, the additional transcripts per million (TPM) normalisation method is often also used.[74]

1.5.5 Differential Gene Expression

The key to differential gene expression (DGE) analysis is the comparison of gene expression across samples. A significant problem with the above methods is that they normalise away the sequencing depth of the experiments which can differ between the samples. The above methods rely on normalising total counts under the assumption of a homogeneous transcript distribution. They can therefore perform poorly should any sample contain a heterogeneous distribution of transcripts, as highly expressed transcripts can skew the distribution of counts.[74–76] All of the newer DGE analysis suites utilised here offer their own forms of normalisation, and all ignore features that are either highly expressed or highly variable when calculating normalisation factors. The following DGE analysis software suites aim to successfully deal with differing library composition.

[†]With single-ended reads, both RPKM and FPKM are equivalent. As paired-end experiments generate two reads that correspond to a single fragment, the latter of the two was created as a way of preventing doubly counted reads per transcript.

empirical analysis of digital gene expression data in R (EdgeR)

One of the more popular DGE analysis suites is EdgeR, from Robinson *et al.*[77] EdgeR utilises the trimmed mean of the M-values (TMM) approach towards normalisation, whereby a single sample is selected as a reference. It selects this sample by creating a library size scale for each gene for each sample, and calculating the average 75^{th} quantile across all samples. The sample with the 75^{th} quantile closest to the average value becomes the reference sample for downstream fold change calculations. Geometric means of fold changes (biased genes) and absolute expression levels (highly or lowly expressed) are calculated from this raw count before being trimmed twice by thresholds defined by these values. In other words, the most "extreme" genes are removed before calculating the weighted scaling factors for each individual gene.

differential gene expression analysis based on the negative binomial distribution (DESeq) & DESeq2

Both DESeq and DESeq2 take the approach of finding a ratio of each raw read count to the geometric mean of all read counts for that specific gene, across all samples. Here the denominator becomes the pseudo-reference sample by which all samples are scaled by. As in EdgeR, by using the geometric mean extreme outliers have less impact in the resulting scaling factors.[78,79]

A series of independent studies have shown the choice of DGE analysis approach to have the potential to markedly affect the outcome of the data, and that no specific package used is likely to yield favourable results for all experiments.[80–82] For this reason, a great deal of care is employed for selecting an apropriate DGE analysis approach for this body of work.

1.6 Statement of Intent and Aims

It is therefore the aim of this body of work to use next-generation sequencing to profile the transcriptome of a genetic model of human hypertension. By identifying biomarkers of the disease, it may be possible to develop a prognostic test for identifying those at risk and to further treat the individual prior to the onset of the disease, mitigating the organ damage that may otherwise have occured. Using this undirected approach may allow for the detection of novel transcript biomarkers that have previously not been implicated in hypertension. It further seeks to do so in peripheral blood, a tissue easily accessible in a clinical setting. Once a robust biomarker is revealed, efforts will be directed towards further characterising its potential involvement in the disease in the hope of revealing novel mechanistic insights into the aetiology of hypertension.

Chapter 2

Candidate Selection Using RNAseq

2.1 Introduction

In order to begin the search for potential biomarkers of hypertension, RNAsequencing can provide a key overview of transcriptomic changes between normotensive and hypertensive rat strains. This high-throughput approach allows a comparison to be made between the WKY and SHR transcripts; profiling their respective abundances and enabling a highlighting of any significant differences between the two. Furthermore, sequencing the transcriptomes of these animals at an age prior to hypertension (within the SHR) will identify transcripts that show the potential to prognostically differentiate between animals that will later develop hypertension and those that will not.

One of the key challenges with primary research involving animal models is how translatable these models are to human biology. This is indeed the main test in taking biological insights gained from the laboratory to the clinic. It is therefore pertinent to conduct a literature review of potential genetic traits that have already been identified within a human population. Only a handful of studies have covered the genetic traits associated with blood pressure, although these lacked replication or were based on limited sample sizes.[83–86] However, one group in particular curated six independent studies and conducted an association assessment to produce a meta-analysis of gene

expression signatures of blood pressure and hypertension.[87] The study made use of patients who were not receiving any anti-hypertensive treatments in order to avoid the possibility of transcriptomic profiles representing pharmaceutical intervention. This allowed the analysis of patterns in the data that were reflective of inherent genetic predispositions. From the 7,017 participants, differentially expressed BP genes were identified from the largest cohort which were cross referenced with the other cohorts to produce a final list of genes implicated in regulating blood pressure. The study continued to perform a meta-analysis of all participants for BP associated genes. This was followed by a GWAS and integrating expressed quantitative trait-loci (eQTL) data in order to differentiate transcriptomic changes from BP from putative causal pathways involved in BP regulation. These final lists enabled a directed approach to be taken with a previously obtained RNAseq output dataset and give rise to the possibility of any downstream biomarkers being translated to a human cohort.

In this chapter the initial approach towards assessing differentially expressed transcripts between a normotensive and hypertensive rat model is outlined, making use of a high-throughput RNAseq analysis approach.

2.2 Aims

Using a previously sequenced n=3 generated dataset, the following will be performed;

- A full QC assessment of previously generated RNAseq raw reads
- Pipeline designed for RNAseq to produce list of differentially expressed genes
- Gene ontology of resulting data
- Candidate selection based on the output data

- Validation using the newly extracted RNA
- Assessment of the false discovery rate

2.3 Materials and Methods

2.3.1 Animals

All animal work complied with the Home Office animals (scientific procedures) act 1986 (ASPA). spontaneously hypertensive rat (SHR) and their age matched Wistar Kyoto (WKY) controls were supplied at either the prehypertensive 3 weeks of age, or the fully hypertensive 12 weeks of age (n=3/group for RNAseq cohort, n=12/group for qPCR validation cohort; Harlan UK Ltd. - Loughborough, UK). Animals were houses in Plexiglas cages in groups of 3 per cage. Each rat had access to food and water *ad libitum* and were supplied with cardboard tubing for cage enrichment. They were maintained in these conditions and checked daily for signs of poor health for 7 days to allow them to habituate to their new environment and being handled. This minimised any potentially confounding effects on our data by way of elevated stress or ill-health.

2.3.2 Sample Preparation

Rats were sacrificed via guillotine, and whole blood was collected from the decapitated rat via a funnel into a potassium ethylenediaminetetraacetic acid (EDTA) tube (BD Biosciences - Franklin Lakes, N.J., U.S.A) to prevent coagulation. The body was gently massaged to prevent internal coagulation. Each blood sample had $\sim 600\mu$l of RNAlater Solution (Ambion Inc. - Austin, Texas, USA) added and was then stored at -20°C.

2.3.3 Blood RNA Extraction

All RNA extraction took place on wet ice and in an RNAse free environment, ensured by good laboratory practice in addition to liberal cleaning of equipment and work spaces with RNAaseZap (Ambion Inc. - Austin, Texas, USA).

Blood samples were initially thawed on wet ice after having been stored at -80°C. The RiboPure Blood Kit (Ambion Inc. - Austin, Texas, USA) was used to extract total blood RNA. All centrifugation steps were conducted at $\sim$ 16,000 x g at room temperature as per the RiboPure Blood Kit protocol. As samples were stored in RNAlater Solution they were first centrifuged for 1 minute. The supernatants was removed and 800μl Lysis Solution and 50μl Sodium Acetate Solution was added to the pellets. After vigorously vortexing to lyse the blood cells, 500μl of Acid-Phenol: Chloroform was added to each sample for the phase separation step. Samples were left for 5 minutes at room temperature before centrifuging for 1 minute to separate the aqueous and organic phases. The upper (aqueous) phase containing the RNA was aspirated to a new tube. 600μl 95-100% ethanol (EtOH) was added to each, and samples were vortexed briefly to ensure mixing. Samples were then passed through a filter cartridge housed in a 2ml tube 700μl at a time. 700μl of Wash Solution 1 was added and washed through the column by centrifuging for 5-10 seconds. Again, flow-through was discarded before 2 x 700μl washes with Wash Solution 2/3 were conducted as centrifugations of 5-10 seconds each. Flow-through was discarded and samples were centrifuged again for 1 minute to remove residual fluid from within the filter. Finally, to elute the RNA the spin cartridges were transferred to a new RNAse free tube and 50μl of Elution Solution (at 75°C) was added to the centre of the filter. Samples were left for 20 seconds before centrifuging for 20-30 seconds to elute RNA. This step was repeated resulting in 100μl of RNA containing water.

2.3.4 RNA Quality Control & cDNA Synthesis

Ensuring samples were constantly placed in wet ice, extracted RNA was assessed for purity and quantity via spectrophotometry (NanoDrop 2000c, ThermoScientific). Once output values were collated samples that presented poor 260:230 and 260:280 ratios were omitted. The finally selected samples were then used for cDNA conversion. The cDNA synthesis kit requires an RNA input volume of 12μl. Since a dilution of samples to a normalised concentration requires fewer potentially damaging steps than a concentrating of samples, samples were normalised to that of the smallest sample concentration multiplied by 12. Samples could then be diluted to this value by adding RNAse free Millipore H_20 up to a total volume of 12μl. This resulted in all samples having the same final concentration prior to cDNA synthesis. cDNA conversion was conducted using the Quantitect RT kit (Qiagen - Hilden, Germany). Normalised samples were incubated with 2μl of genomic DNA wipe-out reagent at 42°C for 2 minutes before returning to wet ice in order to digest unwanted genomic DNA. An Reverse Transcription (RT) mastermix sufficient for all samples (+10%) was then prepared (Table 2.1). 6μl of this mastermix was added to each of the normalised samples before incubation at 42°C for 45 minutes. After, the incubation samples were immediately placed in a water bath set at 95°C for 3mins in order to stop the conversion and denature all of the previously active enzymes. Samples were then rapidly returned to wet ice ready for qPCR analysis.

Reagent	Volume Per 12μl Sample of RNA
Reverse Transcriptase	*1μl*
RT Buffer (5x)	*4μl*
RT Primer Mix	*1μl*
Total Volume	*6μl*

Table 2.1: Table of reagents required for cDNA synthesis Master Mix

2.3.5 RNA Sequencing

For RNA-sequencing, RNA samples were sent to a private sequencing service for additional quality control, library preparation, and sequencing (SourceBioscience Limited - Nottingham, U.K.). Amplified cDNA libraries were prepared from isolated RNA samples previously obtained within the laboratory from 3 animals per strain and sequenced using the Illumina HiSeq 2500 Sequencer (Illumina, Inc. - San Diego, CA, USA). Total RNA samples were enriched by hybridisation to bead-bound rRNA probes using Ribo Zero Kit to obtain rRNA- depleted samples. This was followed by the construction of Illumina libraries using ScriptSeq v2 (Illumina, Inc. - San Diego, CA, USA) that applies unique barcode adapters. The libraries were assessed for their quality using a Qubit dsDNA High Sensitivity DNA Kit and Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA; Agilent High Sensitivity DNA Kit). This followed with further enrichment and amplification of the libraries by qPCR using KAPA Biosystems Library Quantification Kit, before all samples were normalised to 2 nM. Equal volumes of individual libraries were pooled and run on a MiSeq using MiSeq Reagent Kit v2 (Illumina, Inc. - San Diego, CA, USA) to validate the library clustering efficiency. The libraries were then re-pooled based on the MiSeq demultiplexing results and sequenced on a HiSeq 2500 sequencing platform (Illumina, Inc. - San Diego, CA, USA) and cBot with Ver 3 flow cells and sequencing reagents. Library reads of greater than 30 to 35 million were generated for each individual library. The data were then processed using Real-Time Analysis (RTA) and CASAVA thus providing four sets of compressed FASTQ files per library. FASTQ files were concatenated across lanes to yield two files per sample (Reads 1 and 2 of a paired end sequencing run).

2.3.6 Quality Control and Adapter Trimming

All raw reads were pre-processed for quality assessment using the FastQC package.[67] This measures a series of metrics and provided their output in graphical form, allowing

a bioinformatician to assess the need for adapter removal, quality trimming and size selection before pushing forward with the pipeline. A Phred30 quality cutoff was adopted (99.9% base call accuracy).

2.3.7 Alignment

RNAseq alignment and data analysis were all performed in-house using a high-performance computer; "Hydra". This pipeline made use of bash and python scripting to accept RNAseq post-trimmed data as input, before ultimately producing output tables of differentially expressed transcripts. Paired-end (2x100bp) raw input data was initially aligned with Tophat to the sixth iteration of the Rattus norvegicus reference genome (Rn6).[88] HTseq was used to generate read counts using the ENSEMBL Rnor_6.0.97 annotation for reference.[72]

2.3.8 Normalisation of Reads

Here, raw read counts were supplied as input to the pipeline and each statistical DGE calling software suite applied is own normalisation strategy to deal with library size, transcript length, or detection of counts to transcripts (Table 2.2).

Method	Description	Reference
EdgeR	Negative binomial count distribution; genewise dispersion parameter estimation via conditional maximum likelihood; empirical Bayes shrinkage of dispersion parameter; exact test for p-value computation	Robinson, McCarthy & Smyth (2010)[77]
DESeq	Negative binomial count distribution; local regression modeling of mean and variance parameters	Anders & Huber (2010)[78]
DESeq2	Negative binomial count distribution; generalized linear model; shrinkage estimation of dispersion parameter and fold change	Love, Huber & Anders (2014)[79]

Table 2.2: Description of the core modeling strategies of DGE analysis methods for RNAseq. Adapted from Khang, T. F., & Lau, C. Y. (2015)[1]

2.3.9 Differential Gene Expression Predictions

Due to the huge number of available packages for calling differential expression, this pipeline initially made use of several statistical methods from the R Bioconductor suite: DESeq, DESeq2, and EdgeR. Each of these methods allows for the prediction DGE with high confidence, and to utilise the predictions with low p-values in downstream validation. As such, the resulting dataset produced contained all of these prediction approaches to enable an assessment of the differences between them.

2.3.10 Candidate Selection

Primary Candidate List

Over 12,000 transcript elements were flagged up from this comparison, including; pseudogenes, miRNAs and rRNAs, in addition to protein coding transcripts. Using datasets for both blood samples, reads were initially ranked based on EdgeR. The selection of EdgeR to rank by fold change was an arbitrary one, and based on it being established in the literature and of a lower stringency to the other statistical tests. This resulted in a greater list to use for downstream filtering. Unidentified transcripts without a putative role/classification were then discarded, as well as reads identified as non-protein coding so focus could be given to characterised mRNA transcripts. Once these parameters were established, an arbitrary selection of genes whose reads were observed to be most significantly modulated between the WKY and SHR strains gave a preliminary candidate list for validation (Table A.2).

Final Candidate List

From the primary candidate list, 12 transcripts were selected. Half of these were genes that were selected based on EdgeR P-values <0.05 (with the exception of Cat, which had an EdgeR P-value of 0.08 and was included in order to test the accuracy of the DGE software). The remaining 6 transcripts that made up the final candidate list, were

transcripts that were implicated in Huan *et al.*'s 2015 Metaanalysis of hypertensive signature genes and similarly detected within the RN6 WKY vs SHR dataset, regardless of whether or not they were significant or not between the two strains. This resulted in a final list of 12 transcripts that could be used to both validate the RNAseq predicted fold changes observed, and identify potential biomarkers that could discern between the hypertensive strain and its control (Table 2.7).

2.3.11 Oligonucleotide Primers

Once targets were selected, nucleotide sequences of genes of interest were obtained from the NCBI GenBank database. The output FASTA format sequences could be transferred to the NCBI Primer design tool online.[89] Of the options available only; PCR Product Size, Primer Melting Temperatures (T_m), and Organism were altered (Table 2.3). This was in order to yield primers that were more specific to their intended targets, by allowing for a smaller PCR product size and a more narrow range of temperature for primer melting. All other parameters were kept at default settings. Final primers were selected to be complimentary to exonic regions in order to detect both heteronuclear and mature RNA transcripts. Each primer was ran on a single plate to look for relative changes in detection.

Parameter	Default Setting	Amended Setting
PCR Product Size	*Min; 70, Max; 1000*	*Min; 70, Max; 150*
Primer Melting Temperatures (T_m)	Min; 57.0, Opt; 60.0, Max; 63.0, Max Tm Difference; 3	Min; 58.0, Opt; 59.0, Max; 60.0, Max Tm Difference; 2
Organism	*Homo sapiens*	*Rattus novegicus (taxid: 10116)*

Table 2.3: Amended qPCR Primer Design Parameters. Here, the core settings amended include; the expected product size, the primer melting temperature (T_m), and the organism for which the primers are being designed against.

2.3.12 Primer Validation

For conducting any of the RT-qPCR plates, 2μl of cDNA converted from the blood RNA extraction was pipetted out (in duplicate) into a 0.1mL 96-well Reaction Plate (Applied Biosystems, Life Technologies™ - USA) placed on wet ice. A master mix was prepared containing the; forward and reverse primers for the region of interest and the SYBR Green Master mix buffer (Roche - Basel, Switzerland) containing; SYBR Green dye, DNA Polymerase, dNTPs and a salt buffer (Table 2.4). This was made up to a total reaction volume of 13μl per well with RNAse free H_20 (Table 2.5). Also included were a No Template Control (NTC), which included all of the aspects of the reaction mixture outlined above except for the 2μl sample template, and a Millipore H_20 only; both in duplicate. These enabled an assessment as to whether any contamination in the reaction mixture or water would yield any confounding fluorescence that would influence the final C_T reads.

Plates were centrifuged in a desktop centrifuge (PerfectSpin Plate Centrifuge; VWR PeqLab, UK to collect the reaction mix at the base of each well and remove any bubbles that could otherwise optically interfere with the instrument. The plates were then loaded into the StepOne RealTime PCR System (Applied Biosystems, Life Technologies™ - USA). Settings were kept consistent for all qPCR experiments performed, with total reaction volume amended to 13μl, and a run of 40 cycles. The threshold for calculating C_T values was set automatically but checked visually to confirm it was bisecting the exponential phase of the detection curve. A melt curve was performed on each plate after the extension phase of the final qPCR cycle. StepOne RealTime PCR System software automatically rendered a graph displaying changes in fluorescence reporter (Δ-R) against increasing temperature (Δ°C). This enabled the confirmation that no signals were produced due to artefacts such as primer dimerization (Figure 2.1). Once both the C_T and Melt curve graphs were checked, raw C_T values were exported for further use in Microsoft Excel.

Gene Probe	Primer Sequence	
	Forward	Reverse
capza1	*5'-GCTCGTGTGGATGAGTACCT-3'*	*5'-GATTTTGCAGCCAGCACTGA-3'*
ifit1	*5'-GTCACCTTCCTCTGGCTACC-3'*	*5'-ATGGCCTGATGTGCCAATTC-3'*
ankrd35	*5'-AGGTTGGATGGAGCGAAGAT-3'*	*5'-TCTTTGGTGCTCCGTCTCTT-3'*
gstt3	*5'-GCTCGTGTGGATGAGTACCT-3'*	*5'-GATTTTGCAGCCAGCACTGA-3'*
cat	*5'-TATCTCCTATTGGGTTCCCGC-3'*	*5'-GCTGTGCTGACTCCTCTACT-3'*
zcchc9	*5'-TGCGGAGAAATGGGACATCT-3'*	*5'-AAAATGTTCCACGGAGCCAC-3'*
myadm	*5'-TTCTTGGCAGAGATTCCCGA-3'*	*5'-ATTTGCTTACACCCCACCCA-3'*
dusp1	*5'-CTCGGCCAATTGTCCTAACC-3'*	*5'-GAACGCAGAGATCCCAGACA-3'*
glrx5	*5'-CGGAGCTGAGGCAAGGTATTA-3'*	*5'-TCCCACAGATGTGTACTTGCT-3'*
gramd1a	*5'-GGCTTCGTGTATCCTCAGAGA-3'*	*5'-AGTAGCCCACGAGCATCTTT-3'*
ppp1r15a	*5'-TGTCAGAATGCAGAGGCTGA-3'*	*5'-AGTGCACCTTTCTACCCTTCA-3'*
slc31a2	*5'-AAGCCAAGTTGCTCCACAAG-3'*	*5'-TAGGGTTTGTACCTGAGGCG-3'*
eef2k	*5'-CATTGGCCAGTGTTTGGTGA-3'*	*5'-TCATCCAGGTCACTTCGCTT-3'*
tuba1c	*5'-TGCCTTTGTGCACTGGTATG-3'*	*5'-TTCAGCACTATCTGCCCCAA-3'*
eno1	*5'-AGGGTGTCTCAAAGGCTGTT-3'*	*5'-ATTCTCTGTGCCGTCCATCT-3'*
β-actin	*5'-CAGCCGCGAGTACAACCTTC-3'*	*5'-CCCATACCCACCATCACACC-3'*
gapdh	*5'-ATGATTCTACCCACGGCAAG-3'*	*5'-CTGGAAGATGGTGATGGGTT-3'*
rpl19	*5'-GCGTCTGCAGCCATGAGTA-3'*	*5'-TGGCATTGGCGATTTCGTTG-3'*

Table 2.4: qPCR Primer Sequences used for Primer Validation. Primers were designed using the NCBI Primer design tool online with settings amended to those seen in Table 2.3.[89]

Reagent	Volume Per Reaction
SYBR Green Master mix buffer	$6\mu l$
Forward Primer (100ng/μl)	$0.048\mu l$
Reverse Primer (100ng/μl)	$0.048\mu l$
RNAse Free H_20	$4.904\mu l$
Sample cDNA	$2\mu l$

Table 2.5: Table of reagents required for qPCR Master Mix

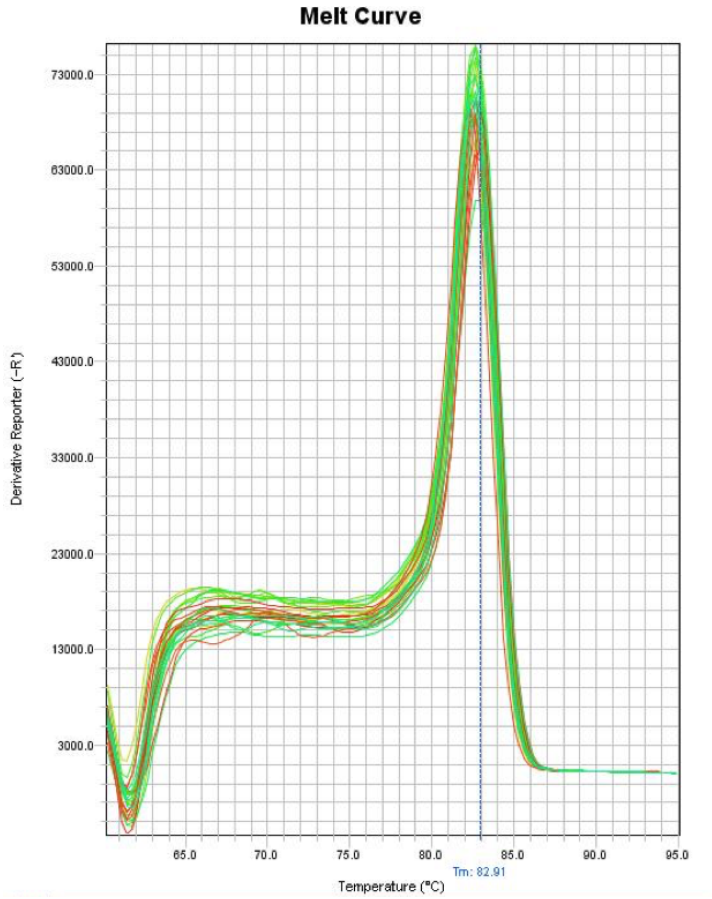


Figure 2.1: Representative plot of a post-PCR melt curve as generated by StepOne RealTime PCR Software (Applied Biosystems, Life Technologies™ - USA). Plot shows the levels of fluorescence against changing temperature to ensure primers show a uniform melting point that would not yield false positive detections.

Once raw C_T values were in Excel technical replicates were first assessed for variation. The Grubb's, or Extreme Studentized Deviate (E.S.D.) test, on GraphPad's Outlier Calculator was used to assess whether or not a replicate was an outlier within a treatment group.[90] The test calculates a specific Z-score for each sample and predicts an outlier if the sample's Z-score is over a threshold specific to the N of the samples supplied. Technical replicates that satisfied this criteria were averaged to produce an average C_T value per biological replicate.

In order to confirm the qPCR primers adhere a linear and predictable increase in fluorescence, proportional to increases in target cDNA, every primer was first subjected

to a calibration run. This was conducted using a pool of all sample's cDNA in order to cover the dynamic range of gene expression for all of the following experiments. A series dilution was made using this pooled cDNA and RNAse free Millipore H_20 to the following dilutions; Neat, 1:2, 1:5, 1:10, 1:20, and 1:50. 2μl of each dilution was used in duplicate to produce a series of average C_T values per dilution.

The following formula was then used to assess primer efficiency and plot this on a standard curve (Figure 2.2, Figure 2.4) at a range of dilutions used;

$$y = kx + d$$

Where; y = Average C_T for a given dilution
k = Line of best fit, using "Least Squares" method
$d = y$ intercept at $Log_e(1)$

Figure 2.2: Equation to represent the linear regression of $\log_n[cDNA]$ *versus* Primer C_T Value.

From this, a PCR efficiency metric was calculated using the equation depicted in Figure 2.3. PCR primers were deemed to be acceptably efficient if this value was between 80-120%. As there were several candidate genes selected for further validation, any genes whose primers failed to show efficiencies within this range were dropped from future analysis. This enabled focus to be given to only the genes with the ability to detect with a higher level of confidence.

$$E = (10^{(-\frac{1}{Slope})} - 1) \times 100$$

Where; E = Primer Efficiency (%)
$Slope$ = Slope coefficient, as denoted by k in equation 2.2

Figure 2.3: Equation to calculate Primer efficiency from linear regression calculated from equation 2.2 (%).

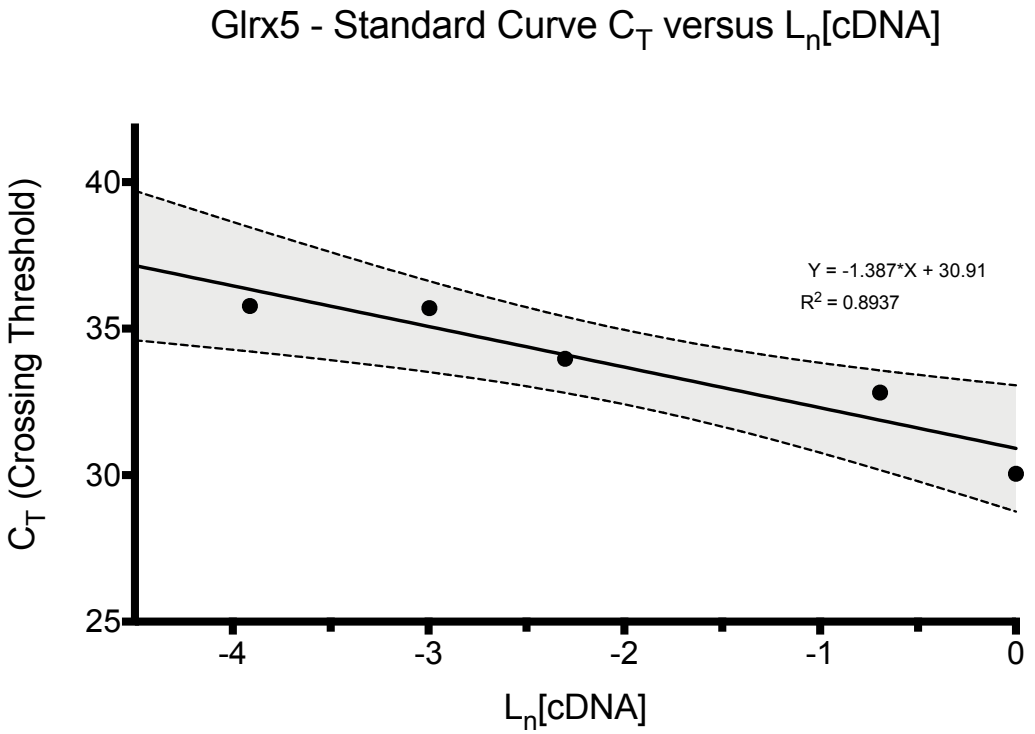


Figure 2.4: Example Standard Curve for assessing primer efficiency (Glrx5). Detection values are plotted against the natural log transformed range of cDNA template dilutions. From the resulting regression analysis, the calculation of the slope was input into the formula outlined below.

2.3.13 qPCR of Target Genes

PCR plates were conducted via the above methodology once primers were deemed to be suitably efficient. An initial plate was ran with β-Actin as the reference gene for blood on all WKY and SHR samples, both juveniles and adults. Afterward, each primer was ran using the same template layout (Figure 2.5) for ease of comparison later on.

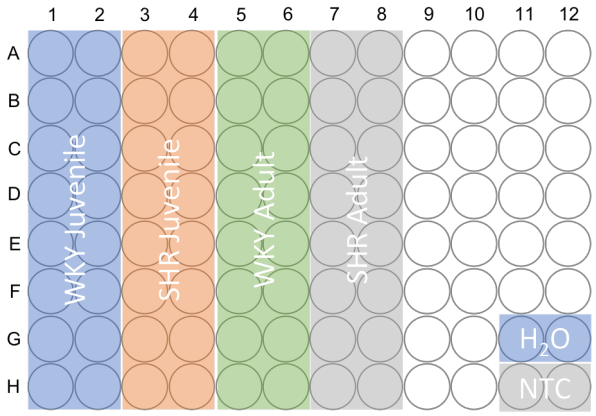


Figure 2.5: qPCR Layout for WKY vs SHR comparisons at both Juvenile and Adult ages

Here, the $2^{-\Delta\Delta C_T}$ method of normalisation was used as per Livak *et al*, 2001.[91] Firstly, absolute C_T values for the appropriate tissue reference gene were selected from the C_T values for the gene of interest to produce a ΔC_T value. Following this calculation, the mean ΔC_T was subtracted from each of the individual sample values giving a $\Delta\Delta C_T$ value per sample. Finally the negative exponent was taken of these values to produce a relative gene expression fold change ($2^{-\Delta\Delta C_T}$). This resulted in a series of values normalised against the selected housekeeping gene, that could be compared relatively against other values on the same plate.

Following a Shapiro-Wilk test for normality, statistics were calculated by conducting an unpaired two-tailed Student's T-test analyses via the GraphPad Prism statistical software package. Genes whose expression did not pass the Shapiro-Wilk test for normality were compared betweem strains using a Mann-Whitney test for significance (Ankrd35, Dusp1, Myadm, and Zcchc9). A confidence interval of 95% (0.05) was

adopted for the threshold of significance. Bar plots were produced using GraphPad and edited for clarity.

2.4 Results

2.4.1 RNAseq Dataset

QC Of RNAseq Data

FastQC output of reads, both pre- and post adapter trimming, showed consistently high mean Phred quality scores across the entire read length (Mean Phred<30, Figure 2.6). Furthermore, per base N content also showed a pass as the mean count for all positions across the reads was below the <5% cutoff threshold. The metric for per sequence quality scores also shows a pass for both pre- and post trimmed reads as the most frequently observed mean quality is the Phred cutoff of >27. The analysis of GC content was conducted to assess potential levels of contamination by assessing the distribution of GC bases across the reads. Here it flagged a failure across all samples as it did not follow the pattern of a normal distribution as expected. The software issues a failure flag if the sum of the deviations from the normal distribution represents more than 30% of the reads. When assessing for over-represented sequences, all of the reads fail the FastQC analysis as each of the read files were found to contain sequences that represent more than 1% of the total reads.

Alignment to a Reference Genome

Tophat produced alignment rates were relatively low for each sample with high levels of discordant or unaligned reads reported (Figure 2.7). On average, 67% of reads were aligned to the reference genome, of which an average of 0.5% were multiple alignments. Furthermore, of the aligned and singly-mapped reads, an average of only 31% were

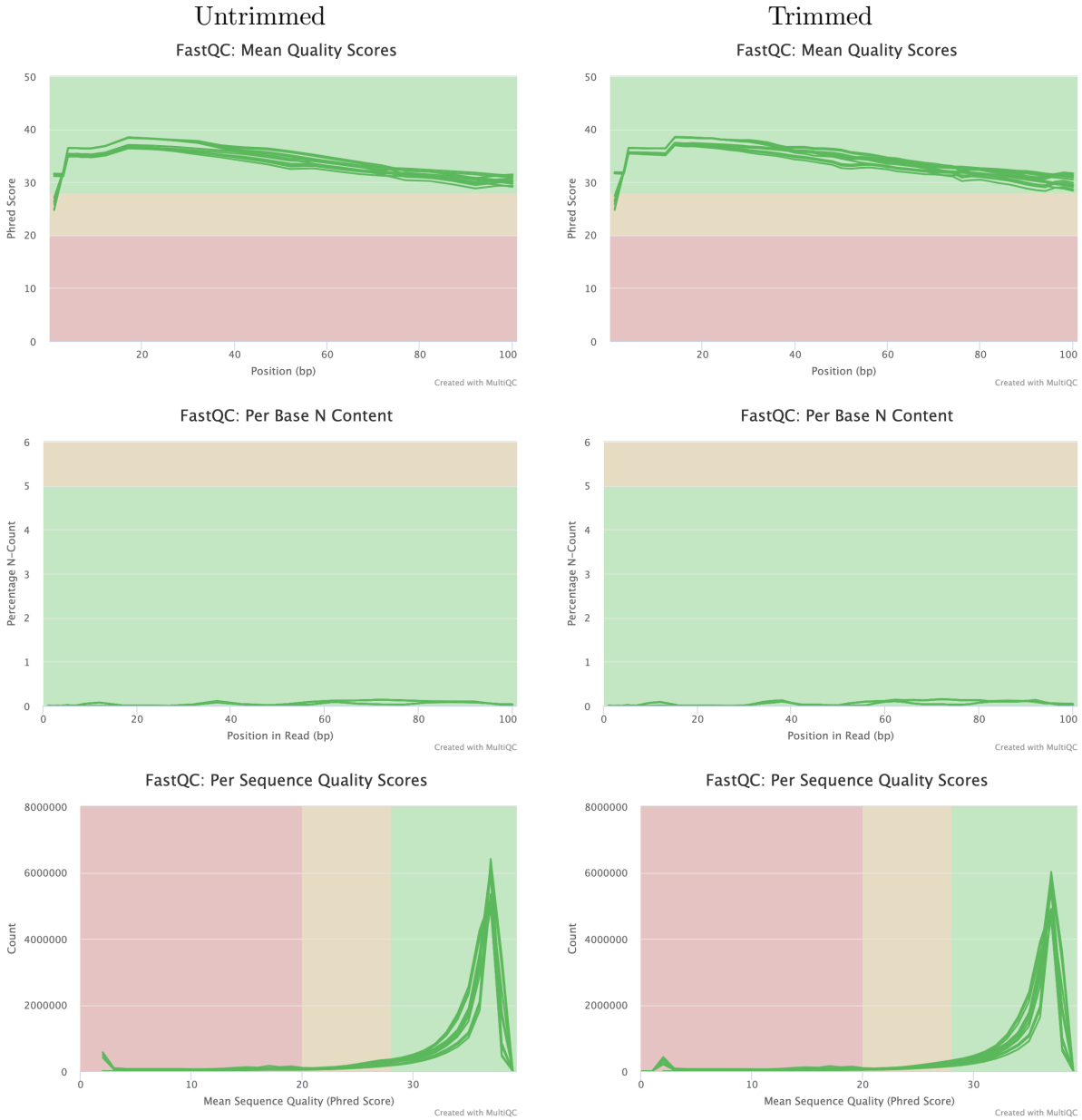


Figure 2.6: RNAseq Quality Control Metrics, as generated by the FastQC package.[67] Here, Quality control metrics are displayed both before and after adapter trimming.

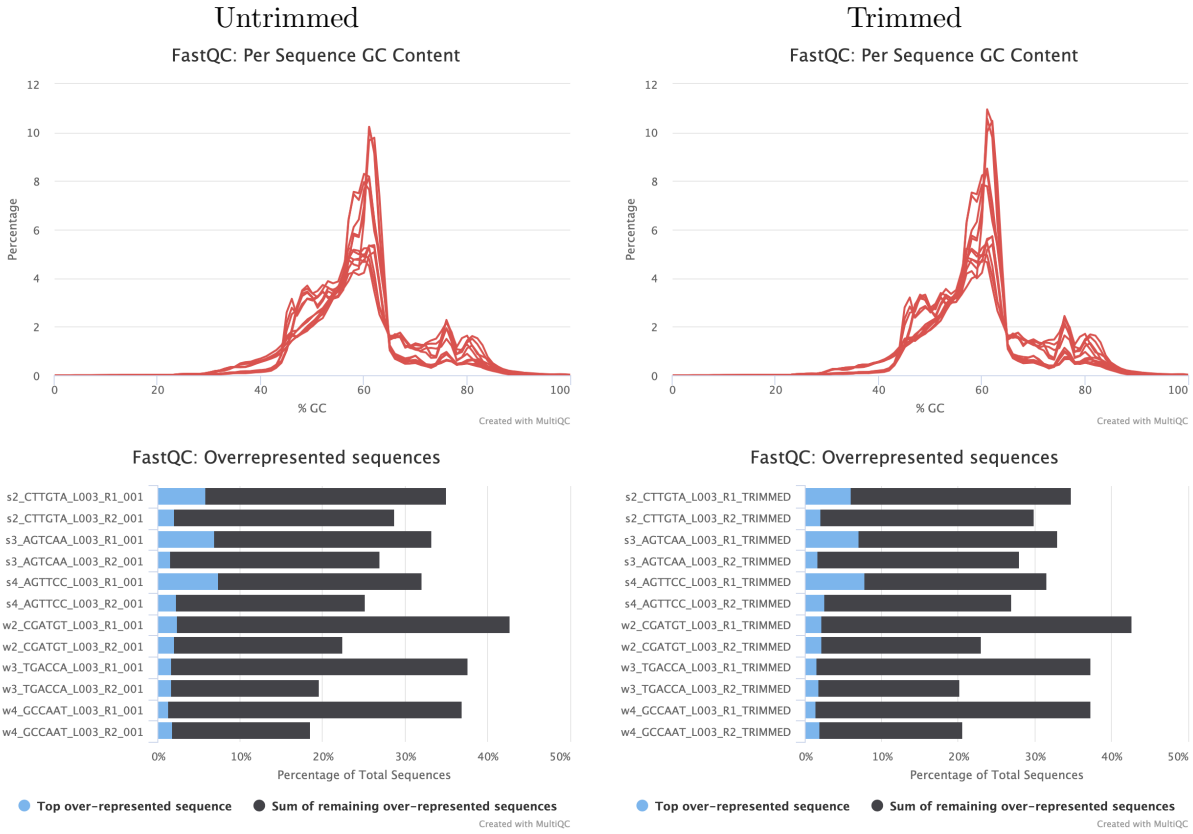


Figure 2.6: RNAseq Quality Control Metrics, as generated by the FastQC package.[67] Here, Quality control metrics are displayed both before and after adapter trimming.

concordantly mapped. Interestingly, the SHR samples had a higher rate of; Total Read Alignment, Multiple Alignments, and Concordantly aligned reads, as compared with the WKY samples (Table A.1).

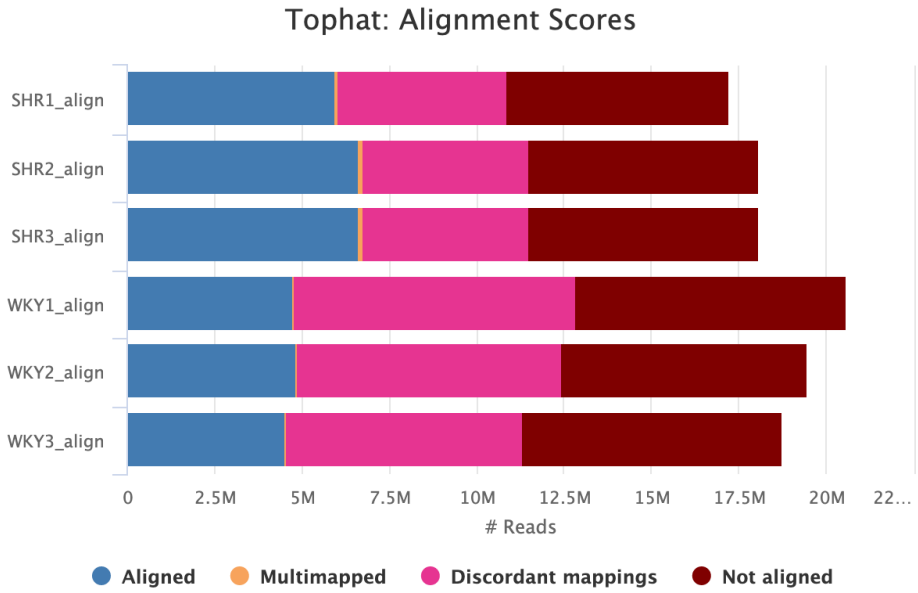


Figure 2.7: Plot of Tophat generated alignment rates (Millions of reads) for the n=3 dataset. Here, the total reads are split into; Aligned, Multimapped, Discordantly Mapped, and Not Aligned.

Generation of Gene Counts

Aligned reads were annotated using the Rn6 Rattus norvegicus genome to produce a count of transcripts for each animal. Running a Principal Component Analysis (PCA) on the samples before and after EdgeR normalisation reveals inherent biases within the data (Figure 2.8). Raw reads reveal a distinct pattern on both PC axis, with two out of three SHR samples forming an isolated cluster on the right hand side of the graph. This separation represents the greatest source of variation (PC1=57.15%). The second PC for the raw reads effectively stratifies between the experimental groups. Post-normalisation, PC1 now appears to resolve the two experimental groups across the X-axis (PC1=77.89%). The PC2 variation no longer isolates samples effectively, and is

now likely representative of biological variation between samples. Here, the necessity for normalisation to prevent RNAseq induced biases from altering our downstream biological interpretation becomes apparent.

Differential Expression

Final predictions for differential expression were curated and the first two pages of "hits", ranked by EdgeR P-Value smallest to highest, are listed in Table A.2. To summarise, 12,564 transcript elements were detected by the RNAseq run, with transcript counts up to 629,055 (Hba-a3, WKY DESeq Generated average count). Adopting a P-Value threshold of <0.05, DESeq predicted 354 transcripts to be significantly regulated between the strains. EdgeR provided a less stringent prediction of 1,821 significantly regulated transcripts and DESeq2 reported 691 transcripts.

Gene ontology analysis of all EdgeR transcripts that had P-Values <0.05 showed the majority of significantly regulated genes to belong to the Hydrolase, Transferase and Oxidoreductase activity GO terms (Figure 2.9).

The finally selected genes are outlined in Table 2.7.

2.4.2 qPCR Data

Validation of Reference Genes for Normalisation

Analysis of putative reference genes in blood revealed a series of significant changes between the strains at both age groups. It should be noted that as there is no reference gene with which to normalise these data, all expression levels are left as unnormalised C_T values, meaning a lower C_T value denotes a higher abundance. Furthermore as these are C_T values, each amplification cycle represents an exponential increase in abundance. Therefore a difference of 1 cycle translates to a doubling or halving of transcript abundance.

β-Actin showed a significant increase in abundance in the SHR within both age

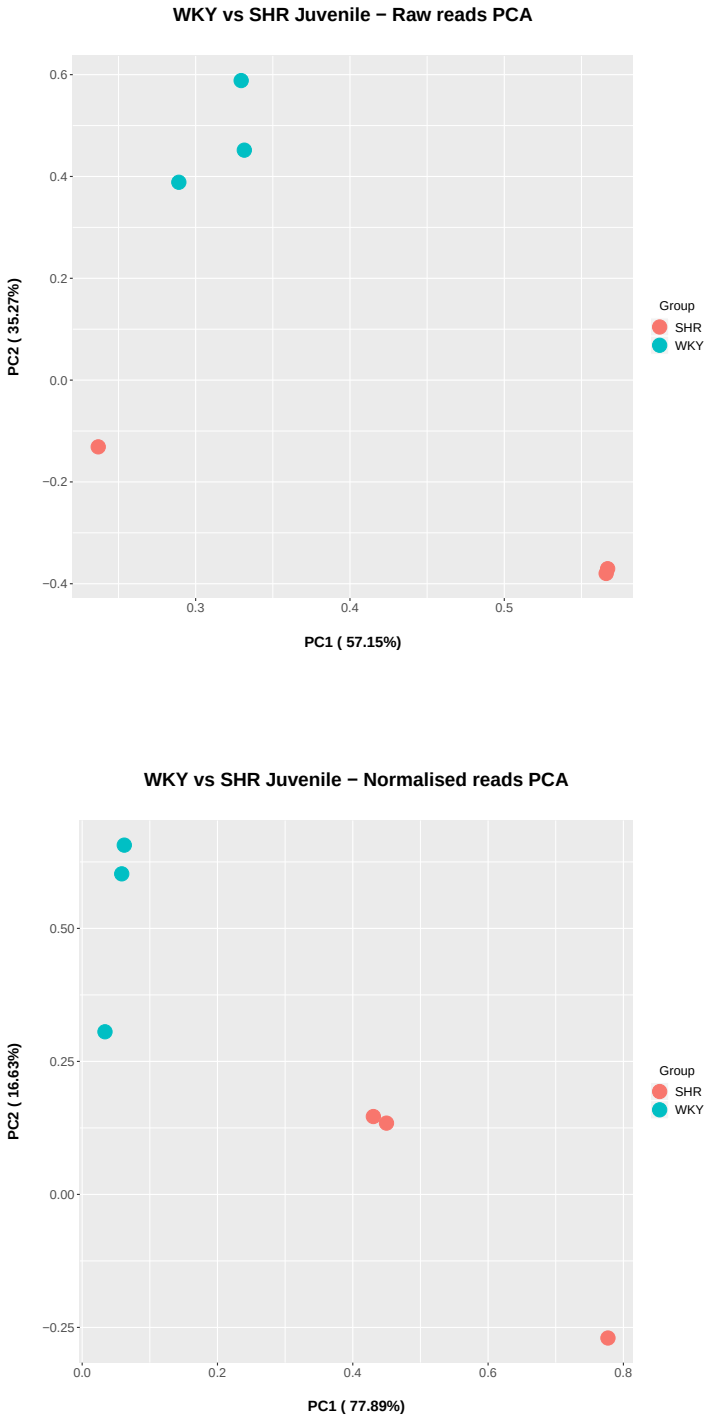


Figure 2.8: Principal Component Analysis (PCA) of n=3 counts. PCA was conducted on both raw (Top) and EdgeR normalised (Bottom) counts. Both PCA outputs reveal a clear separation between the strains.

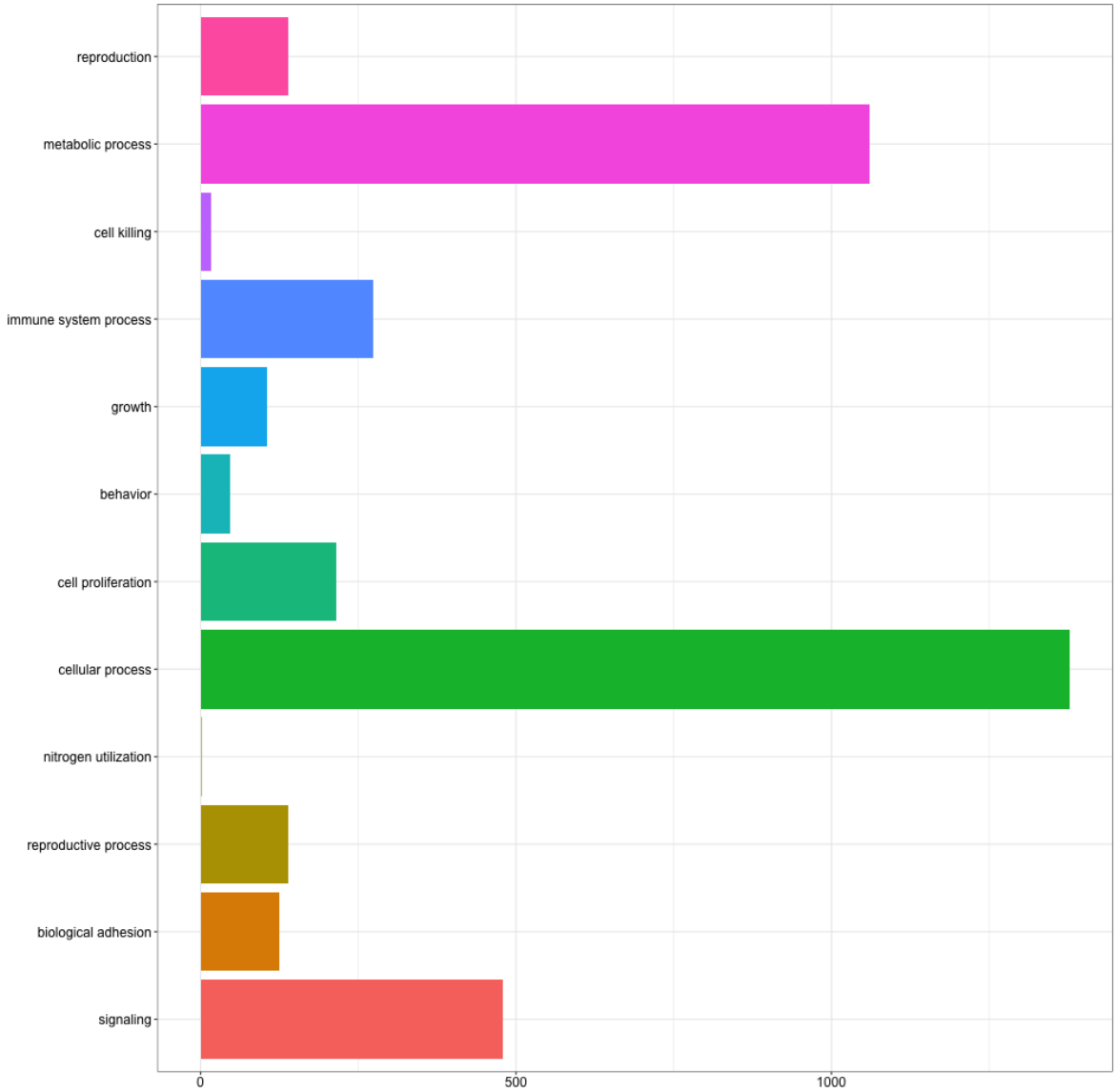


Figure 2.9: Molecular Function Gene Ontology analysis of EdgeR Significant (P-Value<0.05) Transcripts in the Blood.

					DESeq		EdgeR		DESeq2	
		Gene Name	**WKY.Avg**	**SHR.Avg**	**FC**	**P-Value**	**FC**	**P-Value**	**FC**	**P-Value**
Selected	ENSRNOG00000013538	Capza1	85.87	81.55	0.95	1.00	1.00	1.00	0.95	0.92
	ENSRNOG00000019050	Ifit1	296.53	2324.18	7.84	1.30E-45	8.26	1.33E-25	7.47	1.70E-47
	ENSRNOG00000025108	Ankrd35	0.60	91.29	151.25	2.16E-14	121.41	4.42E-23	20.44	3.09E-14
	ENSRNOG00000001242	Gstt3	0.00	132.70	NA	1.60E-22	1464.58	2.78E-28	31.94	2.74E-18
	ENSRNOG00000008364	Cat	11696.90	7572.97	0.65	0.09	0.68	0.01	0.65	1.16E-04
	ENSRNOG00000051682	Zcchc9	2.11	37.30	17.71	5.25E-04	17.93	5.67E-11	7.06	NA
Huan *et al.*	ENSRNOG00000053450	Myadm	43.34	31.92	0.74	1.00	0.77	0.27	0.75	0.51
	ENSRNOG00000003977	Dusp1	41.59	35.33	0.85	1.00	0.88	0.69	0.86	0.79
	ENSRNOG00000004206	Glrx5	4888.10	2072.04	0.42	8.59E-08	0.45	7.55E-07	0.43	7.80E-12
	ENSRNOG00000021106	Gramd1a	48.59	75.08	1.55	1.00	1.60	0.03	1.48	0.23
	ENSRNOG00000020938	Ppp1r15a	594.23	834.08	1.40	0.47	1.47	0.01	1.39	0.01
	ENSRNOG00000013631	Slc31a2	5.54	2.16	0.39	1.00	0.42	0.39	0.65	NA

Table 2.6: Original RNAseq Output of Finally selected Candidate Transcripts. DESeq generated average counts are displayed for each gene, along side fold changes and P-Value predictions from each differential expression analysis software. *N.B.* For DESeq2, the log_2 Fold Change values have been converted to the comparable fold change values. Furthermore, the "NA" P-Values issued by DESeq2 are returned due to a single sample with an extreme outlier (as detected by Cook's distance). Genes are separated on the y-axis to show genes selected from the original candidate list, and those implicated in Huan *et al.*'s meta analysis on hypertension biomarkers.[87]

		Gene	Full Gene Name	Gene Description (Rat Genome Database)
Selected	ENSRNOG00000013538	Capza1	capping actin protein of muscle Z-line subunit alpha 1	Predicted to have actin filament binding activity. Predicted to be involved in actin cytoskeleton organization; barbed-end actin filament capping; and cell junction assembly.
	ENSRNOG00000019050	Ifit1	Interferon-induced protein with tetratricopeptide repeats	Predicted to be involved in cellular response to cytokine stimulus; defense response to virus; and response to bacterium. Orthologous to several human genes including IFIT1B (interferon induced protein with tetratricopeptide repeats 1B).
	ENSRNOG00000025108	Ankrd35	Ankyrin Repeat Domain 35	Orthologous to human ANKRD35 (ankyrin repeat domain 35)
	ENSRNOG00000001242	Gstt3	Glutathione S-transferase theta-3	Predicted to have glutathione transferase activity. Predicted to be involved in glutathione metabolic process.
	ENSRNOG00000008364	Cat	Catalase	Exhibits catalase activity. Involved in several processes, including hydrogen peroxide catabolic process; response to ozone; and response to vitamin. Localizes to several cellular components, including the Golgi apparatus; mitochondrial intermembrane space; and peroxisome. Used to study bacterial pneumonia; brain ischemia; cholestasis; colitis; and lung disease. Biomarker of several diseases, including artery disease (multiple); auditory system disease (multiple); end stage renal failure; eye disease (multiple); and neurodegenerative disease (multiple). Human ortholog(s) of this gene implicated in several diseases, including acatalasia; eye disease (multiple); lung disease (multiple); osteonecrosis; and pseudoxanthoma elasticum. Orthologous to human CAT (catalase).
	ENSRNOG00000051682	Zcchc9	Zinc Finger CCHC-Type Containing 9	Predicted to have nucleic acid binding activity and zinc ion binding activity. Predicted to be involved in negative regulation of phosphatase activity. Orthologous to human ZCCHC9 (zinc finger CCHC-type containing 9).
Huan *et al.*	ENSRNOG00000053450	Myadm	Myeloid Associated Differentiation Marker	Predicted to be involved in several processes, including negative regulation of heterotypic cell-cell adhesion; negative regulation of macromolecule metabolic process; and negative regulation of protein kinase C signaling. Orthologous to human MYADM (myeloid associated differentiation marker).
	ENSRNOG00000003977	Dusp1	Dual Specificity Phosphatase 1	Exhibits growth factor binding activity and protein tyrosine/threonine phosphatase activity. Involved in several processes, including negative regulation of nitrogen compound metabolic process; regulation of apoptotic process; and response to testosterone. Localizes to the nucleus. Used to study Huntington's disease. Biomarker of several diseases, including anti-basement membrane glomerulonephritis; brain disease (multiple); hypertension; pancreatitis; and portal hypertension. Human ortholog(s) of this gene implicated in breast cancer. Orthologous to human DUSP1 (dual specificity phosphatase 1).
	ENSRNOG00000004206	Glrx5	Glutaredoxin 5	Predicted to have several functions, including 2 iron, 2 sulfur cluster binding activity; electron transfer activity; and protein disulfide oxidoreductase activity. Predicted to be involved in protein lipoylation. Localizes to the dendrite; neuronal cell body; and nucleus. Human ortholog(s) of this gene implicated in autosomal recessive pyridoxine-refractory sideroblastic anemia 3. Orthologous to human GLRX5 (glutaredoxin 5).
	ENSRNOG00000021106	Gramd1a	GRAM Domain Containing 1A	Predicted to have cholesterol binding activity and cholesterol transfer activity. Predicted to be involved in cellular response to cholesterol. Orthologous to human GRAMD1A (GRAM domain containing 1A).
	ENSRNOG00000020938	Ppp1r15a	Protein Phosphatase 1 Regulatory Subunit 15A	Exhibits protein phosphatase 1 binding activity. Involved in several processes, including cellular response to nutrient levels; cellular response to organonitrogen compound; and regulation of cellular protein metabolic process. Localizes to the cytosol. Biomarker of brain ischemia and transient cerebral ischemia. Orthologous to human PPP1R15A (protein phosphatase 1 regulatory subunit 15A).
	ENSRNOG00000013631	Slc31a2	Solute Carrier Family 31 Member 2	Predicted to have copper ion transmembrane transporter activity. Predicted to be involved in cellular copper ion homeostasis and regulation of copper ion transmembrane transport. Orthologous to human SLC31A2 (solute carrier family 31 member 2).

Table 2.7: Additional information on candidate transcripts. Displayed here are the ensembl gene ID accession numbers alongside their associated gene name. Also displayed are the full gene names and descriptions; sourced from the Rat Genome Database.[92]

groups, at around 1 fold difference which approximately translates to a doubling in transcript abundance (Figure 2.10; juvenile WKY vs SHR, -0.79 fold, *P-Value=0.032*; adult WKY vs SHR, -1.01 fold, *P-Value=0.038*).

A similar, but more significant difference was observed for GAPDH as SHR blood again saw an increase in its expression as compared with WKY blood (Figure 2.10; juvenile WKY vs SHR, -0.94 fold, *P-Value=0.022*; adult WKY vs SHR, -0.98 fold, *P-Value=0.023*).

RPL19 showed the greatest change in abundance between the strains, at both ages. This was in the same direction as the other putative reference genes tested, but saw a greater magnitude of change (Figure 2.10; juvenile WKY vs SHR, -1.04 fold, *P-Value=0.008*; adult WKY vs SHR, -0.84 fold, *P-Value<0.0001*).

Producing the Geometric Mean, by use of the BestKeeper software suite, did little to normalise the data towards consistent C_T values across the groups. The C_T values still saw a significant decrease when comparing SHR to WKY (Figure 2.10; juvenile WKY vs SHR, -0.92 fold, *P-Value=0.013*; adult WKY vs SHR, -0.907 fold, *P-Value=0.0045*).

Biomarker Candidates

qPCR analysis on my selected transcripts revealed a series of changes taking place within the whole blood. Expression of Capza1 in the whole blood showed now significant differences between the strains at either age group (Figure 2.11a; juvenile WKY vs SHR, +0.1 fold, *P-Value=0.60*; adult WKY vs SHR, -0.1 fold, *P-Value=0.55*). Furthermore there was a much greater spread within the juvenile animals than in the adult animals, of either strain.

Expression levels of MYADM showed a significantly increased level in the SHR juvenile as compared with the WKY. However this difference was not reflected in the adult cohort (Figure 2.11a; juvenile WKY vs SHR, +0.36 fold, *P-Value=0.04*; adult WKY vs SHR, +0.12 fold, *P-Value=0.11*).

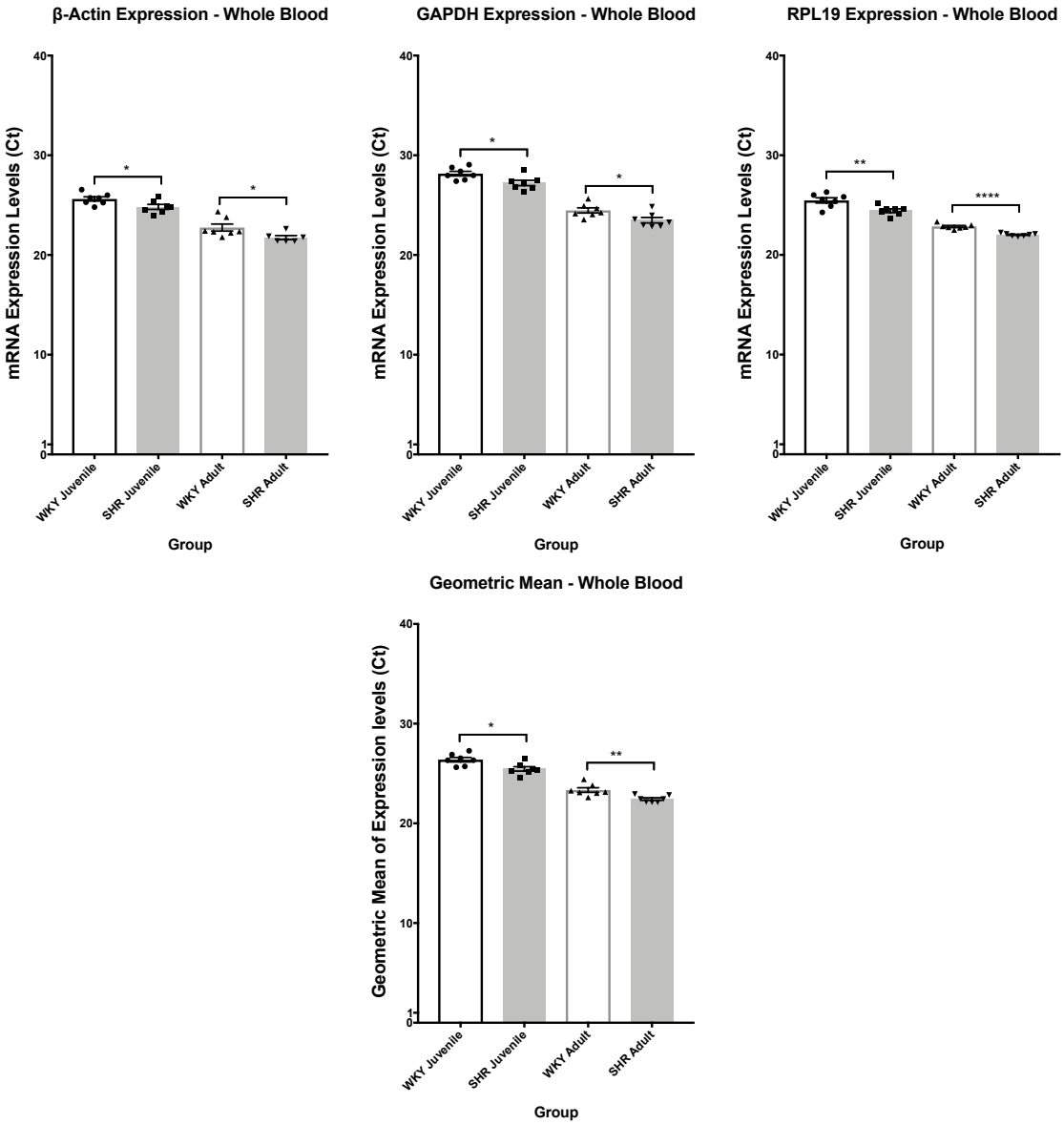


Figure 2.10: qPCR ΔC_T values for putative reference genes; β-Actin, Glyceraldehyde 3-phosphate dehydrogenase (GAPDH), and 60S Ribosomal Protein L19 (RPL19). Statistics are presented as unpaired Student's T-Test P-values, between strains at any given age (P-value<0.05, *; P-value<0.01, **; P-value<0.001, ***).

DUSP1 expression saw a significant decrease across the juvenile comparison. However this was not replicated within the adult comparison. The control WKY juvenile group showed a much greater level of variation between each animal compared with the other groups whose expression levels were more stable (Figure 2.11a; juvenile WKY vs SHR, -0.75 fold, *P-Value=0.0051*; adult WKY vs SHR, -0.001 fold, *P-Value=0.88*).

Transcript levels of Glrx5 were elevated in the juvenile cohort as compared with the adult cohort, however no significance existed between the strains at either age group (Figure 2.11a; juvenile WKY vs SHR, +0.01 fold, *P-Value=0.96*; adult WKY vs SHR, +0.01 fold, *P-Value=0.71*).

Gramd1a showed a significant increase in transcript abundance in the juvenile SHR compared with the juvenile WKY. This change was not observed across the strains in the adult comparison (Figure 2.11a; juvenile WKY vs SHR, +0.61 fold, *P-Value=0.003*; adult WKY vs SHR, +0.02 fold, *P-Value=0.90*).

Neither age group showed a significant change in expression levels for Ppp1r15a in the blood (Figure 2.11a; juvenile WKY vs SHR, +0.1 fold, *P-Value=0.64*; adult WKY vs SHR, -0.01 fold, *P-Value=0.83*).

Ifit1 saw an incredibly significant and robust elevation in transcript abundance in the SHR as compared with the WKY. This was consistent in both the juvenile and adult age groups. However the increase was much more pronounced within the juvenile cohort (Figure 2.11b; juvenile WKY vs SHR, +68.2 fold, *P-Value<0.0001*; adult WKY vs SHR, +8.31 fold, *P-Value<0.0001*).

qPCR showed Ankrd35 levels to be significantly regulated between the strains, with SHR blood showing much higher levels of the transcript as compared with their age-matched controls. This increase in abundance was greater in magnitude within the juvenile cohorts, however was more tightly grouped within the adult comparison (Figure 2.11b; juvenile WKY vs SHR, +3.31 fold, *P-Value=0.0005*; adult WKY vs SHR, +0.544 fold, *P-Value=0.0002*).

SLC31A2 expression in the blood was not significantly different between either age

group comparison. A slight increase in spread was observed within the SHR juvenile sample as compared with the WKY juvenile group (Figure 2.11b; juvenile WKY vs SHR, -0.13 fold, *P-Value=0.44*; adult WKY vs SHR, -0.05 fold, *P-Value=0.29*).

Expression of Gstt3 in whole blood showed a robust and significant elevation in the SHR vs the WKY in both age groups. Furthermore all groups showed a tight clustering of expression levels with the exception of the WKY juvenile group, where expression of Gstt3 appeared to greatly vary (Figure 2.11b; juvenile WKY vs SHR, +43.4 fold, *P-Value<0.0001*; adult WKY vs SHR, +14.8 fold, *P-Value<0.0001*).

Levels of Cat remained relatively unchanged, with no significance reported from the unpaired Student's T-Test. Variation was much greater within the juvenile animals as compared with the adult animals, reaching an almost significant increase in the juvenile SHR cohort (Figure 2.11b; juvenile WKY vs SHR, +0.68 fold, *P-Value=0.09*; adult WKY vs SHR, +0.03 fold, *P-Value=0.57*).

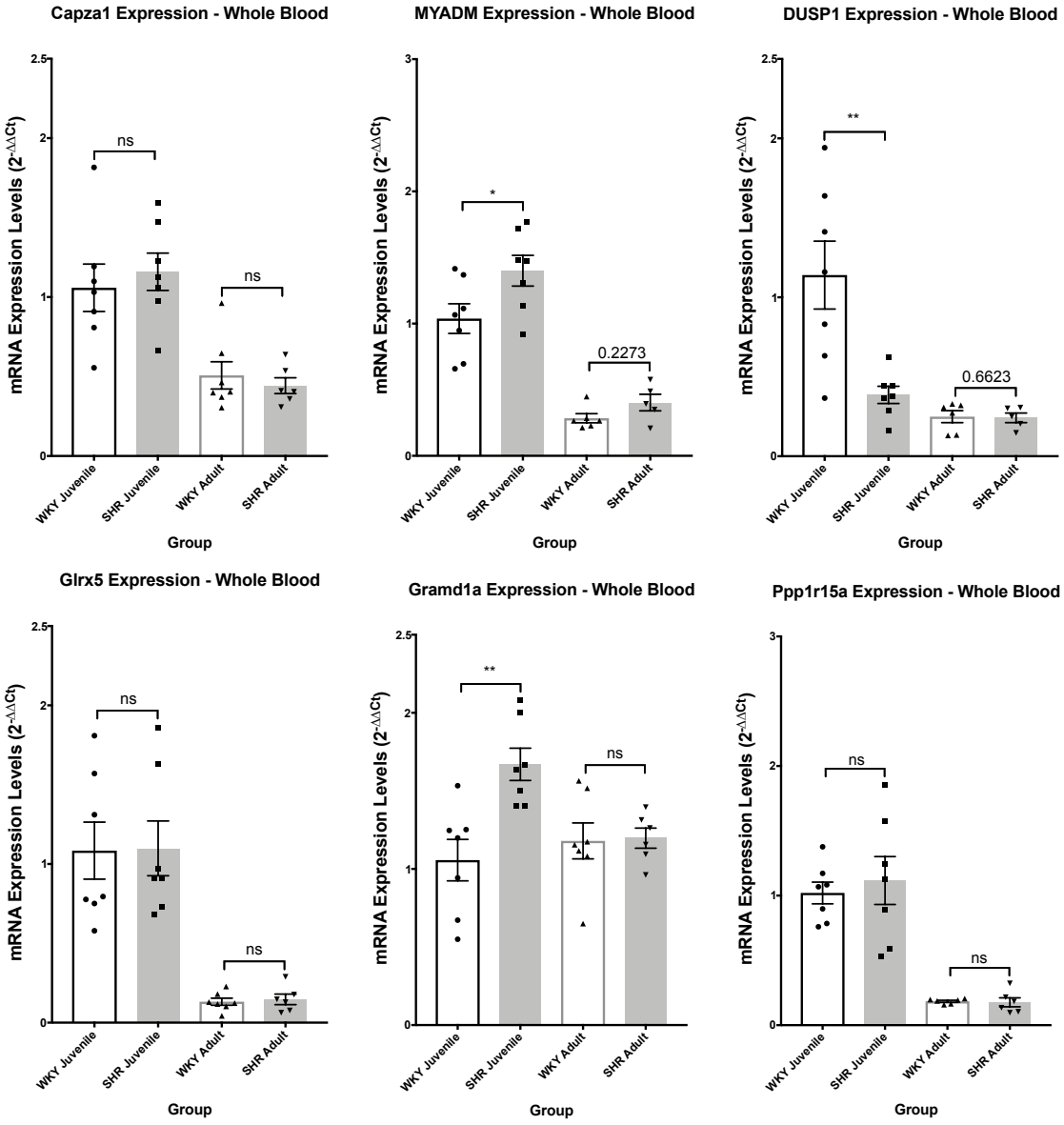
Zcchc9 showed a robust increase in the SHR vs the WKY across both age groups. This increase was much greater in magnitude in the juvenile comparison, compared with the adult comparison (Figure 2.11b; juvenile WKY vs SHR, +5.89 fold, *P-Value<0.0001*; adult WKY vs SHR, +0.47 fold, *P-Value<0.0001*).

2.5 Discussion

2.5.1 RNAseq Data

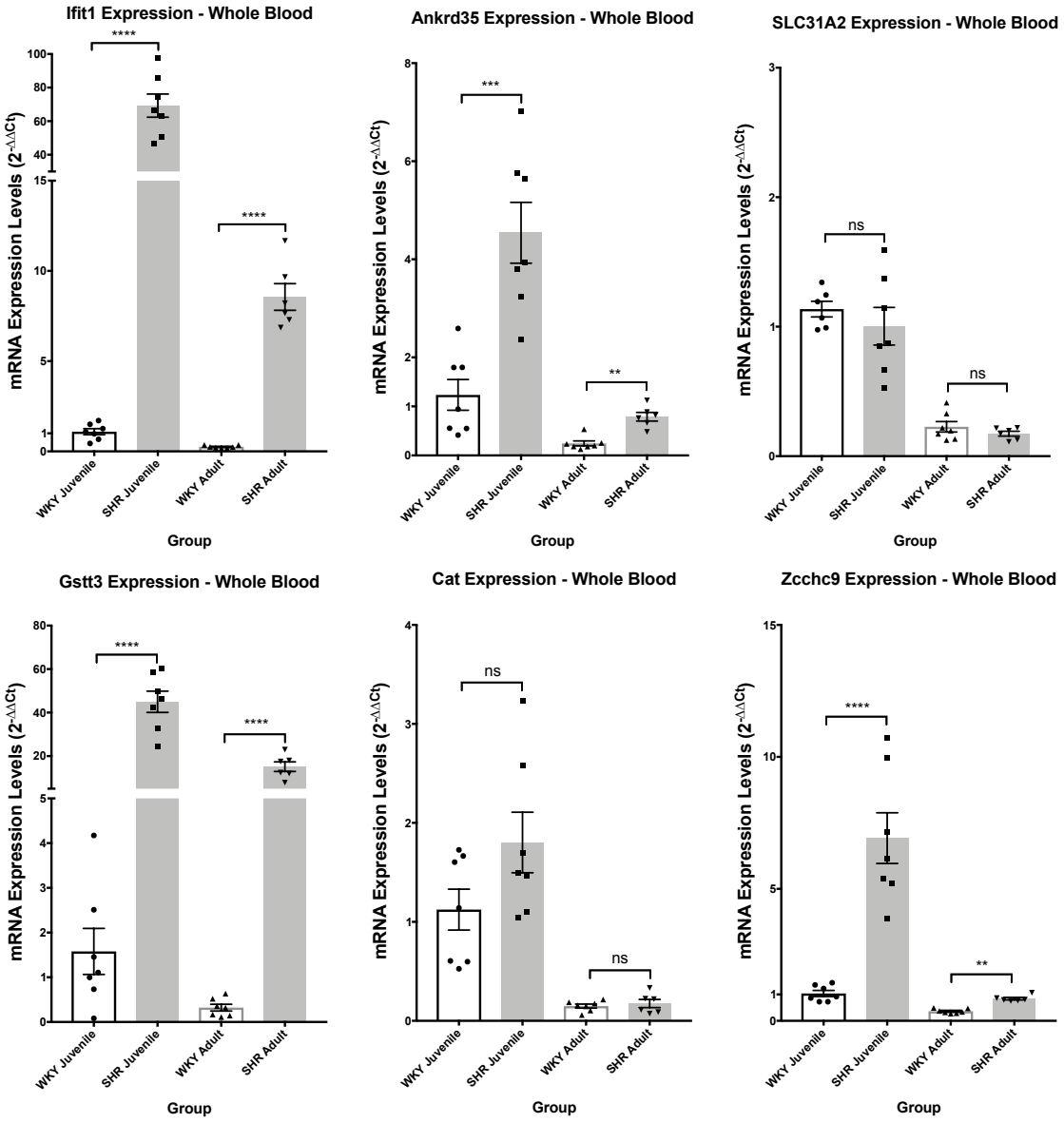
Potential issues to QC Output

The first step following an RNAseq experiment, is to assess the quality of the resulting reads. From this QC step many of the raw .fastq read files flagged up as failing the various QC metrics FastQC assesses. While mean quality scores appeared consistently high (Phred Score>30) across the entire reads and per sequence, as well as the all



(a)

Figure 2.11: Relative mRNA expression of Candidate Biomarkers in the whole blood of Wistar Kyoto (WKY) and Spontaneously Hypertensive Rats (SHR) at both 4- and 12 weeks of age. Here, expression is presented as $2^{-\Delta\Delta C_T}$, normalised to relative expression of β-Actin, using the WKY as the control group for each normalisation. Statistics are presented as unpaired Student's T-Test P-values, between strains at any given age (P-value<0.05, *; P-value<0.01, **; P-value<0.001, ***).



(b)

Figure 2.11: Relative mRNA expression of Candidate Biomarkers in the whole blood of Wistar Kyoto (WKY) and Spontaneously Hypertensive Rats (SHR) at both 4- and 12 weeks of age. Here, expression is presented as $2^{-\Delta\Delta C_T}$, normalised to relative expression of β-Actin, using the WKY as the control group for each normalisation. Statistics are presented as unpaired Student's T-Test P-values, between strains at any given age (P-value<0.05, *; P-value<0.01, **; P-value<0.001, ***).

reads having a low percentage N-count, the GC Content flagged up as a fail across all samples. Warnings to this module indicate a problem with the library itself and may have been resultant of either a specific contaminant (adapter dimers for example) or a contamination with a different species.[93] Many of the potential issues that arose are dealt with by trimming adapters that are left on from the indexing stage of the library preparation. However, the trimming of reads failed to improve upon the GC content.

2.5.2 qPCR Data

Having selected a series of potential biomarker candidates, qPCR provided a different experimental approach to validate both their biological validity in another cohort of animals, in addition to the RNAseq generated dataset itself.

One of the problems faced in any validation study is that of identifying a suitable housekeeping gene for the comparison. Quantitative RNA data are relative to both the total quantity and constituent transcripts contained within the extracted RNA. Therefore the reference genes used for standardization are of critical importance, and any analysis of the validity of qPCR data must consider the relevance of the genes selected for their ability to successfully normalise any discrepancies in the input RNA.

Initially RPL19 was investigated to normalise gene expression against. This was due to its previous usage as a reference gene by the laboratory within brain tissues and therefore primers were readily available.[94] qPCR analysis of blood tissue showed significantly increased levels of RPL19 in the SHR compared with the WKY across both age groups. For this reason a suitable housekeeping gene that remained stable between the strains in blood tissue needed to be identified. Multiple studies have investigated the stability of reference genes in different tissues, but very few have done so in blood. It was for this reason that the putative reference genes; β-Actin and GAPDH were investigated. Both were significantly modulated across the strains and were therefore not suitable to normalise against. The Bestkeeper software produced

by Pfaffl *et al* (2004) allowed the coordination of all of the reference genes analysed to give a standardised geometric mean to compare the candidate genes against.[95] This approach resulted in a set of housekeeper values that were still significantly modulated in blood tissue across strains for both age groups. Moreover, the P-values obtained between the strains were more significant than those obtained for β-Actin alone, and so β-Actin was selected as the reference gene due to its comparably low significance. While this may not impact on the analysis of profoundly modulated gene expression, such as that observed for ifit1, it is of crucial importance to conclude with any confidence that the modulation is resultant of the experimental variable and not merely an inherent difference across the experimental groups. Further efforts must be made in order to identify a stable housekeeper for future validation steps. This could be achieved by using more putative housekeeping genes to produce a more stable geometric mean. An alternative strategy could be to consult the RNAseq data to investigate the genes that were not significantly modulated between the strains. Once properly validated, this could provide a stable housekeeping gene to normalise against in future studies.

2.5.3 Comparison of RNAseq and qPCR output

False Discovery Rate

When comparing the RNAseq changes in gene expression with the qPCR analysis, multiple disparities were observed in either the fold change or level of significance (Table 2.8, Table 2.9). Many of the changes statistically significant in the RNAseq data were not as significant in the qPCR findings and *vice versa*. There are many reasons why this could be the case, all of which should be considered when interpreting these results.

A potential source of variation may have arose from the method by which the libraries were prepared. Here libraries were depleted of rRNA by use of the Ribosomal Depletion method. A comparison of this method against a Poly(A) selection method of library preparation in blood has shown ribosomal depletion to lead to a higher

		DESeq			EdgeR		DESeq2		qPCR	
Gene Name	**WKY.Avg**	**SHR.Avg**	**FC**	**P-Value**	**FC**	**P-Value**	**FC**	**P-Value**	**FC**	**P-Value**
Capza1	85.87	81.55	0.95	1.00	1.00	1.00	0.95	0.92	1.1	0.60
Ifit1	296.53	2324.18	7.84	1.30E-45	8.26	1.33E-25	7.47	1.70E-47	63.51	<0.0001
Ankrd35	0.60	91.29	151.25	2.16E-14	121.41	4.42E-23	20.44	3.09E-14	3.68	0.0005
Gstt3	0.00	132.70	NA	1.60E-22	1464.58	2.78E-28	31.94	2.74E-18	28.47	<0.0001
Cat	11696.90	7572.97	0.65	0.09	0.68	0.01	0.65	1.16E-04	1.6	0.09
Zcchc9	2.11	37.30	17.71	5.25E-04	17.93	5.67E-11	7.06	NA	6.67	<0.0001
Myadm	43.34	31.92	0.74	1.00	0.77	0.27	0.75	0.51	1.35	0.04
Dusp1	41.59	35.33	0.85	1.00	0.88	0.69	0.86	0.79	0.34	0.0051
Glrx5	4888.10	2072.04	0.42	8.59E-08	0.45	7.55E-07	0.43	7.80E-12	1.01	0.96
Gramd1a	48.59	75.08	1.55	1.00	1.60	0.03	1.48	0.23	1.58	0.003
Ppp1r15a	594.23	834.08	1.40	0.47	1.47	0.01	1.39	0.01	1.09	0.64
Slc31a2	5.54	2.16	0.39	1.00	0.42	0.39	0.65	NA	0.96	0.44

Table 2.8: Comparison of RNAseq output with qPCR validation on candidate biomarkers. *N.B.* For DESeq2, the log_2 fold Change values have been converted to the comparable fold change values. Furthermore, the "NA" P-Values issued by DESeq2 are returned due to a single sample with an extreme outlier (as detected by Cook's distance).

DGE Test	No. of False Positives	false discovery rate (FDR)
DESeq	1	0.14%
DESeq2	3	0.43%
EdgeR	2	0.29%

Table 2.9: false discovery rate (FDR) of RNAseq differential gene expression (DGE) tests, as calculated by False Positives (DGE P-Value<0.05 $\wedge$ qPCR P-Value<0.05) / True Positives (qPCR P-Value<0.05).

percentage reads aligning to intronic regions.[96] The study finds that 220% more reads would have to be sequenced in order to achieve the same level of exonic coverage when using rRNA depleted libraries as compared with Poly(A) selected libraries. As the focus of this study was primarily expression of mature transcripts, a Poly(A) selection method of library preparation may optimise the detection of reads aligning to exonic regions; utilising a greater percentage of the available read depth and ultimately improving upon the accuracy of resulting DGE calling.

One of the causes of disparity between the RNAseq output and the qPCR validation may find its roots in the poor QC output from the RNAseq reads, in particular the GC content. FastQC raised a failure flag across all reads for this metric and indicated the source of the error to be either adapter dimer contamination or a contamination from another species. Due to the width of the peaks across the distribution it suggests the latter is the case.[93]

It should also be noted that these discrepancies may well be due to true biological differences between the samples and not a reflection of any bias inherent to the experimental method used to measure transcript expression. While these are both inbred rat strains, there will still exist a level of biological noise at these given sample numbers. It therefore becomes a question of replicates necessary to get the most out of the differential expression analysis. Experiments with a low number of replicates will often have poor statistical power to be able to resolve differentially expressed genes correctly and furthermore cannot account for natural biological noise.[97,98] With this in mind an experimenter could simply increase the number of replicates to improve upon statistical

power and to ultimately improve upon correctly identifying differentially expressed transcripts. However the relatively high cost of RNAseq, coupled the drive for primary researchers to reduce the number of animals used, make this approach prohibitive. For this reason it seems pertinent to tease out the precise relationship between accurately resolving differentially expressed genes and the replicate number required to do so. A few studies have sought to answer this question. Two of which took publicly available RNAseq data from 21 clones of two separative strains of mouse.[80,99,100] Soneson & Delorenzi (2013) compared various DGE tools and measured for a level of concordance between the tools' outputs in order to assess their suitability. They concluded none of the 11 DGE tools they assessed performed acceptably with fewer than 3 replicates.[80] Burden *et al* use the metric of FDR to rank five DGE tools, concluding that at least six replicates are required per condition to achieve a low FDR.[100]

A study in 2016 was the first of its kind to use a much greater dataset of 48 biological replicates across 2 conditions in order to answer this question and provide guidelines for future RNAseq experiments.[101]The study continues to suggest that for future RNAseq experiments, at least 6 replicates per group should be considered with an optimum of 12 replicates per condition where identifying the majority of all DGE is required. Furthermore, they state that for experiments with <12 replicates per condition; EdgeR or DESeq2 is to be used. This is in accordance with studies that provide a much lower level of replication

2.5.4 Concluding remarks

Here 3 biological replicates per strain are relied upon for the RNAseq analysis, which is confirmed by the qPCR validation to be greatly underpowered for detecting a series of robust differentially expressed transcripts. Further optimisation is therefore required in order to lay a firm foundation for additional validation of biomarkers of hypertension.

Chapter 3

Optimisation of High-Throughput approach

3.1 Introduction

3.1.1 Building upon previous findings

Given the high false discovery rate observed for the n=3 dataset, it is pertinent to revisit the standard practice for any given RNAseq experiment. As an RNAseq pipeline presents a long string of individual manipulation events, both pre- and post sequencing, there exist many variables that could potentially lead to confounding results. The vastly different pipelines and methods of library preparation do not lead to directly comparable datasets. Instead, here the focus is given to the main potential sources of error, in the hopes of optimising each step in the approach for future experiments.

3.1.2 *In Silico* Optimisation

Over the last few years, increasing numbers of novel differential expression methods have been produced, each assuming a different statistical spread to best represent the relationship between raw RNAseq counts and actual abundance of genes.

A key issue with assessing the different normalisation strategies is the need to validate each test's predicted expression changes using other technologies in order to account for the presence of False Discovery Rates. Classically this is done by selecting a cross-section of transcript candidates and performing qPCR to assess how accurate the predicted fold changes are when compared to the biologically observed fold changes qPCR shows. While this is the best way to validate the large datasets NGS produce, it is time consuming and can always be improved upon by including more and more validation candidates.

One of the ways in which a pipeline can be optimised is via the inclusion of synthetically generated sequencing data. With a standardised dataset, the number of differentially expressed genes can be set, and furthermore compared with the number genes predicted to be differentially expressed by each statistical method. Here the R studio program CompcodeR is used, which enables the synthesis of a dataset to the specifications set out by the user.[102] The program makes use of the data generation method outlined in Soneson and Delorenzi (2013).[80]

3.1.3 Repetition of RNAseq Dataset

Using the literature, and findings from the *In Silico* optimisation, a repeat of the previous RNAseq experiment was conducted. Based on these results, alterations were made to the number of replicates, library preparation, and the analysis pipeline itself. All of which will be inspected to see how each alteration can serve to optimise the approach, and lead to results more consistent with the qPCR data obtained from a greater number of replicates in the previous chapter.

Library Preparation

One of the first major decisions to be made during an RNAseq experiment, is the way in which the library of transcripts will be made. ribosomal ribonucleic acid

(rRNA) is by far the most highly abundant constituent of the total transcriptome, for any animal or human tissue, comprising 80-90% of the total RNA.[103] Therefore, the transcripts of interest must be isolated in order to get the most from the finite number of reads each RNAseq experiment is capable of. With this in mind, two main options exist to isolate mRNA transcripts; Poly(A) selection, and rRNA depletion. Poly(A) selection involves the selection of polyadenylated RNA (including mRNA transcripts) using oligo(dT) primers to bind to the polyadenylated tails. rRNA depletion takes an alternative approach to removing the highly abundant rRNA transcripts. Using magnetic beads with rRNA complimentary sequences, rRNAs hybridize to the beads and can be subsequently removed from the total RNA sample by isolating the magnetic beads that now contain them. Each of these approaches has a series of unique advantages and disadvantages, with many studies discussing these at great length.[104–107]

Where input is limited, or potentially degraded, rRNA depletion manages to leave fragmented transcripts in the sample library. This is due to the potential lack of polyadenylated tail, preventing these fragments from being enriched by use of the Oligo(dT) selection process.[104,105,107] As the input for this series of experiments is high in abundance, and low in degradation, this does not present a problem that needs overcoming.

With such high quality input tissue, the key consideration for this library preparation is how to maximise the number counts generated via RNAseq, and specifically for those types of transcripts that are the focal point of this study. As touched upon briefly in section 2.5.3, Poly(A) selection presents the best way in which to maximise the sequencing depth as the focus here is on the protein-coding fraction of the transcriptome.[96,104,106]

The previous n=3 dataset was obtained through rRNA depletion (Section 2.3.5). Here the new RNAseq experiment makes use of the Poly(A) selection method of library preparation in an attempt to see if this can resolve the high levels of duplicated reads observed in the n=3 dataset, allowing more reads to be used for transcripts of interest.

Number of replicates

Building upon the *In Silico* optimisation, and the literature review conducted in section 2.5.3, the new RNAseq experiment will make use of an optimised number of replicates in order to minimise the false discovery rate.

Analysis Pipeline

Once samples are resolved by the RNA sequencer, the resulting reads need to mapped to a reference genome in order to identify precisely which genes these original transcripts were produced from. For this reason, the ability to align as many reads as possible to the reference, and to do so accurately, has the potential to translate to significant changes in the ultimate DGE analysis.

The advent of RNAseq technology gave rise to a new challenge when trying align reads to a reference genome, as new software had to take into account alternative splicing. A new generation of splice-aware software was therefore created, with several advances over the earlier BLAST-like alignment tool (BLAT).[108,109] In 2009, Tophat was released as part of this new generation, and made use of a two-step approach to read alignment. Firstly, reads were assessed to discover exon junctions, before using this information as part of the final alignment to the reference genome.[88] This, and similar tools, set the standard in read alignment algorithms for a long time and was therefore a key step in the previous pipeline employed for the n=3 dataset. However, year on year additional programs are released, with each purporting to be capable of; greater alignment rate, higher alignment accuracy, or faster alignment compute speeds. All of which are valuable attributes to any RNAseq experiment.

As with the DGE prediction software, several groups have sought to compare these approaches with both biologically and *in silico* generated datasets. A paper in 2013, compared 26 mapping protocols based on 11 programs and found major differences between these methods based on a series of performance benchmarks.[110] The

benchmarks used included; total alignment yield, exon junction discovery, suitability of alignments for transcript reconstruction, mismatch and gap placement, and basewise accuracy. These tests were conducted on paired-end RNAseq reads from a human leukemia cell line (K562), mouse brain lysates (C57BL/6NJ), as well as two simulated human transcriptomes.[111–113] From the biologically derived samples (i.e. the K562, and CB57BL/6NJ), both were Poly(A) selected libraries. From their comparison, the spliced transcripts alignment to a reference (STAR) software performed the best across a host of the performance metrics tested, most notably the highest level of primary alignments devoid of mismatches, and the highest accuracy in splice-detection.[70] Based on the findings of this paper, and others, STAR was selected as a new splice-aware aligner to compare against the previous Tophat alignment software.[114,115]

3.1.4 Microarray vs RNAseq

While RNAseq presents the most current and thorough way to perform a high-throughput analysis of the transcriptome, microarrays have continued in their popularity over the last few years due to their reliability and relatively low cost. This is especially true when studying a well characterised model organism with a well defined genetic annotation, such as the *Rattus norvegicus*, as microarray chips can yield an enormously comprehensive snapshot of the transcriptomic profile of an animal. Here the focus is on transcripts that are significantly modulated between the two experimental strains being studied. With this in mind, microarray presents an approach that is more than capable of shining light on these differences, and can furthermore act as an independent technology by which to validate the RNAseq based approach used here. Therefore, a comparison between an RNAseq and Microarray output would provide a key foundation in pushing forward with biomarker discovery.

3.2 Aims

Building upon the n=3 dataset, priority is given to assessing potential sources of the high false discovery rate, and optimising the approach for subsequent studies. This will be achieved via the following;

- An *In Silico* assessment of the optimum number of replicates
- An *In Silico* assessment of the optimum statistical software suite for calling differential gene expression
- Performing a microarray of differential gene expression
- A repeat of the RNAseq experiment, with an adjusted library preparation method and number of replicates
- A comparison of microarray output with RNAseq output

3.3 Materials and Methods

3.3.1 *In Silico* Optimisation

Comparison of Methods (*n=3*)

Using the CompcodeR program in an RStudio environment, a series of synthetically generated RNAseq datasets were produced. The *generateSyntheticdata* command includes many different arguments to best replicate the type of data the finally adopted differential expression method is to be used on. The below table (Table 3.1) includes the settings used for the base comparison of methods (n=3; reads = $1x10^7$; 12,500 total genes; 10% differentially expressed). These parameters were set to be comparable to the previous RNAseq data the laboratory had processed. The initial CompcodeR

comparison therefore acted as a benchmark test of methods, before alterations to parameters were made for further optimisation of our pipeline. This run was performed via the *runComparisonGUI* command, and repeated 10 times with all parameters left as default (0.05 for all settings). CompcodeR then runs a comparison of methods based on these 10 replicates.

Argument	Amended Setting	Description
n.vars	12500	Initial number of genes to be generated in the simulated data set.
Samples.per.cond	3	Number of samples in each condition (n).
n.diffexp	1250	Number of genes simulated to be differentially expressed between the included conditions.
seqdepth	1.00E+07	Total number of mapped reads.
Fraction.upregulated	0.5	Fraction of differentially expressed genes that are upregulated
Between.group.diffdisp	FALSE	Sets whether or not levels of dispersion are different between conditions
Filter.threshold.total	1	Threshold for total of any given gene across all replicates.

Table 3.1: List of key arguments to *generateSyntheticdata* command. Here are the specific parameters used for generation of the dataset used for the Comparison of Methods (section 3.3.1). For the Comparison of Replicates (section 3.3.1), the *Samples.per.cond* parameter was adjusted for the generation of each dataset to be used.

Comparison of Replicates

Once a comparison was made for a synthetic dataset with 3 replicates per group, a comparison of varying n values was made by maintaining all parameters as above, but altering the *sample.per.cond argument*. In order to narrow the optimisation, DESeq2 was selected for comparing the number of replicates due to its performance in the n=3 comparison against other DGE software.

3.3.2 Tissue Preparation

Animals were sacrificed, had blood extracted and RNA isolated as outlined in Chapter 2 materials and methods. As blood tissue is primarily comprised of haemoglobin containing erythrocytes for the transportation of oxygen, the blood transcriptome comprises approximately 80-90% globin mRNA. Evidence also shows an abundance of globin gene

transcripts originating from erythrocyte progenitor cells, as globin RNA has been seen to comprise >70% of the transcriptome of adult human reticulocytes.[116] Samples therefore required depleting of these transcripts in order to prevent downstream techniques from only detecting the overly abundant globin.[117] This was achieved by subjecting the RNA samples to a GLOBINClear^{TM} Protocol.[118] Samples were initially assessed for RNA concentration using the Nanodrop 2000c (ThermoFisher) and normalised with RNAse-free water for RNA reactions between 1-10μg and input volumes of $\leq$14μl before running through the GLOBINclear procedure. This made use of biotinylated oligos, complementary to globin mRNA, to hybridize to the globin transcripts. Streptadvidin magnetic beads were then added to bind with the now biotinylated globin transcripts, and a magnetic microcentrifuge rig captured and isolated the bound complexes. This allowed for globin-free total RNA supernatant to be decanted, leaving the vast majority of the globin mRNA behind.

The final RNA samples were sent to a private sequencing service for further analysis including library preparation, microarray and RNAseq assessment (SourceBioscience Limited - Nottingham, U.K.).

3.3.3 N=6 RNAseq

For RNA-sequencing, libraries were built using the TruSeq Stranded mRNA kits (Illumina, Inc. - San Diego, CA, U.S.A.; Figure 3.1). In a similar approach to the globin depletion protocol outlined above, polyadenylated mRNA was first isolated by use of binding complimentary oligo(dT)s bound to magnetic beads. This was contrary to the previously obtained RNAseq dataset (n=3) which made use of rRNA depletion rather than mRNA selection. The magnetic beads were then magnetically isolated from the RNA pool, along with the bound polyadenylated mRNA transcripts. The remaining mRNA was then cleaned, fragmented, and primed ready for cDNA synthesis (Step 1; Figure 3.1). Using random nucleotide primers, the first cDNA strand is initially

synthesised before actinomycin is used to prevent second strand synthesis (Step 2; Figure 3.1). After this, the RNA template is degraded to ensure only the second strand is used for next cDNA synthesis. Here Deoxyuridine Triphosphate (dUTP) nucleotides are used in place of Deoxythymidine Triphospgate (dTTP) nucleotides, in order to allow for a differentiation between the first and second strands of cDNA later on (Step 3; Figure 3.1). The newly synthesised double stranded cDNA is next prepared for adapter ligation via an additional adenylation at the 3' end of each strand (Step 4; Figure 3.1). This enables the cDNA to hybridize with the thymine overhang on the 3' end of the adapters. Adapters are ligated onto the cDNA fragment and dUTPs are enzymatically removed from the fragments (Step 5; Figure 3.1). Following enzymatic removal of the dUTPs, fragments are PCR amplified to produce a resultant library that can be validated quantitatively via qPCR (Step 6/7; Figure 3.1). Additional quality control takes place using a 2100 Bioanalyzer (Agilent Technologies) to assess fragment size and to ensure primer dimerisation has not occurred (Step 8; Figure 3.1). This quality control step is also employed at various stages along the sample preparation to ensure RNA integrity is maintained. Unlike spectrophotometric methods of RNA analysis electrophoresis methods measure the sizes of RNA fragments, enabling information to be gleaned as to how degraded the transcripts are at any given step. Since each sample was ligated to an adapter containing a specific nucleotide sequence, libraries could then be pooled for running on the flow cell (Step 9/10; Figure 3.1).

Once libraries are washed over the flow cell cDNA fragments hybridise, via their adapters, to complimentary olignucleotides bound on to the flow cell. At this point, cluster generation will occur via bridge amplification to yield isolated clusters born from the same original fragment. After clustering, libraries were loaded onto an Illumina HiSeq 4000 flow cell (Illumina, Inc. - San Diego, CA, USA) and ran to obtain a final count of around 52 million paired-end reads per sample at 2 x 75bp read length.

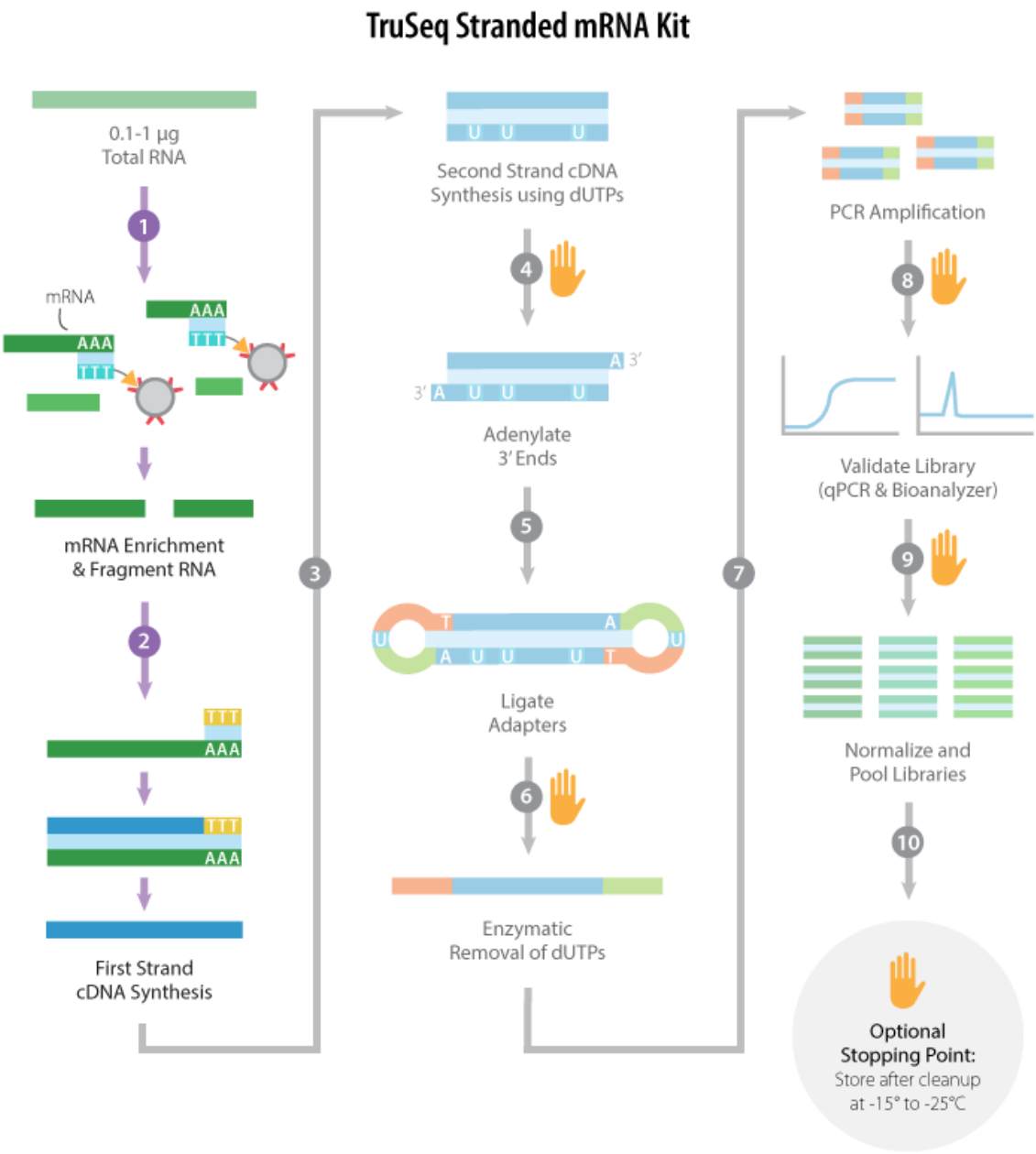


Figure 3.1: Overview of TruSeq Stranded mRNA Preparation protocol. Image adapted from www.abmgood.com

3.3.4 Microarray

For microarray analysis of the whole blood transcriptome, the current market leader for producing assay chips is ThermoFisher Scientific. When studying expression profiles within the rat, ThermoFisher offers two options as part of their Clariom^{TM} range. The Clariom D^{TM} Rat 1.0 Assay was selected over the Clariom S^{TM} due to its ability to perform a comprehensive and detailed analysis of both coding and noncoding genes, in addition to exons and splice variants. This greater level of resolution comes at a greater financial cost, however it does futureproof the data for a more comprehensive analysis if required.

In order to prepare for microarray analysis, the GeneChip^{TM} Whole Transcript Pico protocol was followed (ThermoFisher Scientific,[119]). Isolated and purified RNA was initially incubated with a reverse transcriptase and PCR primers containing a T7 promoter. This constituted the first-strand cDNA synthesis step, and primes the 5' of the transcripts for further manipulation downstream. Next, a 3' adapter is added to the single-strand cDNA, and used as a template for double-stranded cDNA synthesis in a pre-*in vitro* transcription (IVT) amplification reaction. The reaction makes use of a DNA polymerase and RNase H to degrade the RNA and synthesise a single-stranded cDNA containing a 3' adapter. The resulting single-stranded cDNA could then be converted to double-stranded cDNA, enabling it to act as a template for *in vitro* transcription. This was achieved through Taq DNA polymerase and a set of Adapter-specific primers to simultaneously synthesise and pre-amplify the double-stranded cDNA. Once the double-stranded cDNA had been sufficiently amplified, anti-sense RNA, or complimentary RNA (cRNA), could be synthesised and further amplified via an *in vitro* transcription using the double-stranded cDNA as a template and the T7 RNA polymerase. This approach towards RNA preparation is derived from the Eberwine or RT-IVT method, outlined by Van Gelder *et al.*, 1990.[120] The samples were next purified of any enzymes, inorganic phosphates, salts and unutilised nucleotides, using purification beads. This

was in order to prepare the cRNA samples for the second-cycle single-stranded cDNA synthesis.

At that point, a spectrophotometry quality control checkpoint was carried out to assess purity and quantity of the cRNA samples and to normalise libraries ready for continuing. From the normalised libraries of cRNA, a sense-strand cDNA was synthesised through reverse transcription and second-cycle primers. The resulting sense-strand cDNA incorporates dUTP at a maintained ratio to the incorporated dTTP, preparing the cDNA for later fragmentation. RNase H was next used to hydrolyse the cRNA template, leaving the single-stranded cDNA. Following hydrolysis, another purification step was carried out on the second-cycle single-stranded cDNA and, after a second spectrophotometry quality control checkpoint, the samples were ready for cDNA fragmentation and labeling. The purified, sense-strand cDNA was next biochemically fragmented by a combination of Uracil-DNA Glycosylase (UDG) and Apurinic/ayrimidinic Endonuclease 1 (APE1), utilising the unnatural dUTP residues introduced earlier in the protocol, to break the DNA strand. The resulting fragments were next labeled by terminal deoxynucleotidyl transferase (TdT) using ThermoFisher's propriety DNA labeling reagent that is covalently linked with biotin. Samples were then loaded onto the GeneChip for hybridisation via the conjugated biotin molecules. The GeneChip was put through a series of wash and staining steps before being scanned ready for image anaylsis.

Once the raw output data was returned from the sequencing company, it was be put through a series of stringent quality control steps. All further analysis was carried out on a desktop computer, making use of the R environment to run Oligo and Affy packages to process the raw .cel files.[121,122] The first step in quality control assessment, is the generation of chip pseudo-images to shed light on potential inconsistencies on individual arrays (Figure 3.18). A further quality control is to plot the distribution of the log base 2 intensities ($Log_2(\mathrm{PM}ij)$ for array i and probe j) of the perfect match probes, in order to compare probe intensity behaviour across the different arrays (Figure

3.19). The shape and distributions of the intensities should all be relatively consistent across the different arrays, with any discrepancy indicating a potential source of error that requires additional normalisation. Box plots show the same differences in probe intensities across chips, however allow for a clearer the overall distribution than in the histograms; especially when assessing the effects of normalisation.

There are several sources of potentially confounding noise in microarray experiments. There may be inconsistencies with the quantity of RNA used for hybridisation, subtle differences in hybridisation conditions, imperfections on the chip array surface, or an imperfect synthesis of the probes themselves. Any of these sources of noise may lead to the observation of differences between the arrays that are not reflective of biological variability, and therefore need to be removed in order to make accurate conclusions about the data. In order to normalise both within and across each chip, the Robust Multiarray Averaging (RMA) method was used.

The Robust Multiarray Averaging (RMA) method of normalisation makes use of only the Perfect Match (PM) probes, and performs a series of steps. Initially, a background correction is performed in order to correct for spatial variation within each individual array. This is achieved by calculating a background-corrected intensity for each PM probe to ensure all background corrected intensities are positive. It then makes a Log_2 transformation of the background corrected intensities to improve the distribution of the data in both up- and down-regulated directions. The Log_2 transformed data then undergoes a Quantile normalisation in order to correct for sources of noise across the arrays. This makes the expression data comparable between the arrays.

3.4 Results

3.4.1 *In Silico* Optimisation

Comparison of Methods (*n=3*)

Once the software was run, a series of investigative plots were drawn in order to compare how accurate the various statistical tests were for that given synthetic dataset. These plots are displayed below as representative plots of 1 replicate out of 10.

In Figure 3.2 a numerical representation of Receiver Operating Characteristics (ROC) curve is shown. Total area under the curve (AUC) tends towards a value of 1 as the test is better at detecting truly positive genes ahead of truly negative genes. This plot shows the total performance of each test, however indiscriminately towards alterations in true positive stringency. When the total plot is taken into account, EdgeR GLM marginally performs the best at a score of 0.82. This is closely followed by DESeq2 and DESeq GLM at 0.81.

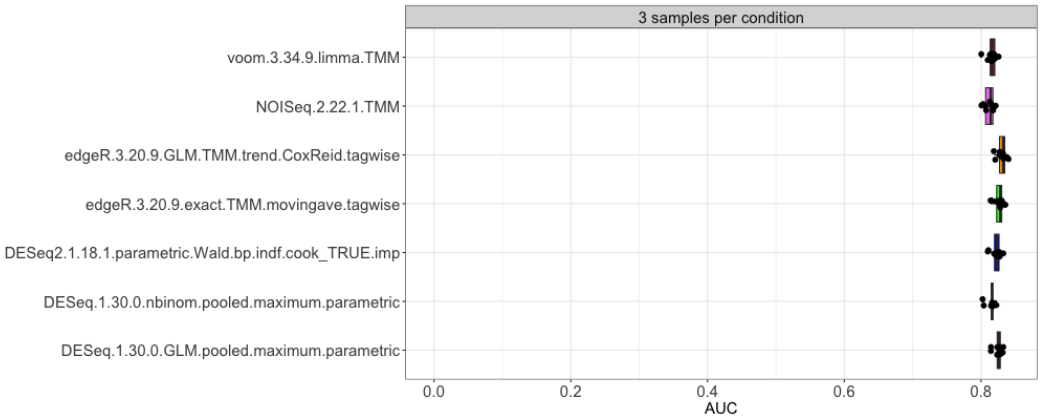


Figure 3.2: Area Under the Curve (AUC)

Figure 3.3 shows a summary of ability of a test to rank truly positive genes (y-axis) ahead of truly negative genes (x-axis). A good test procedure maximises the area under the curve and tends towards the upper left of the plot. Here all tests perform similarly under these conditions. As the true positive rate is low (i.e. threshold for significance is more stringent) DESeq2 and Voom-Limma appear to show the lowest levels of false

positive calling. As the true positive rate increases, however, EdgeR and NOISeq TMM appear to best mitigate false positive calling.

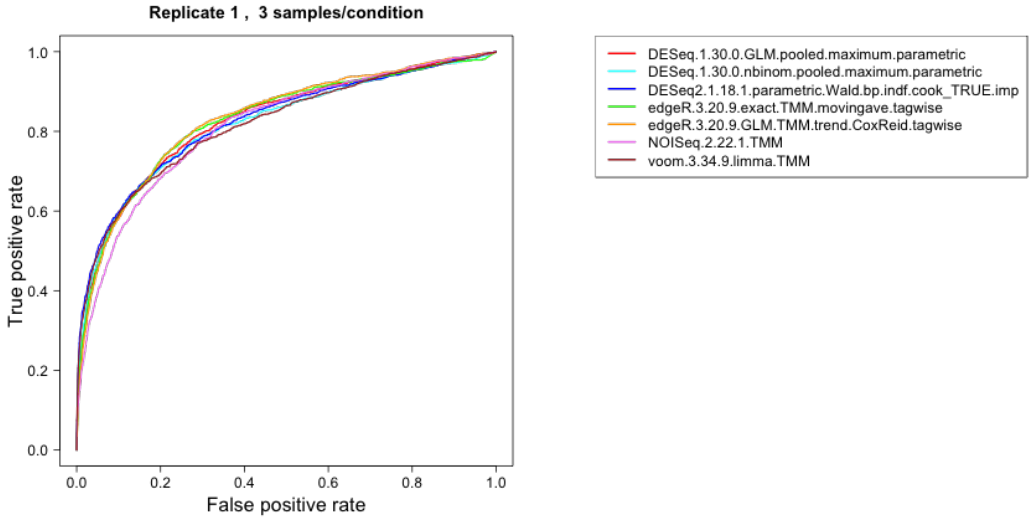


Figure 3.3: ROC Curves

Figure 3.4 represents a visual depiction of the number of false discoveries reported (y-axis) against a list of genes ranked for significance via their respective statistical test (x-axis). Hence, at any given position on the number of selected genes, the y-axis represents the number of truly non-differentially expressed genes that are ranked above that position. A good statistical test places few true negatives amongst the top-ranked genes, and therefore is shown by a curve tending to the bottom right of the plot with a slow rise on the y-axis. Here both DESeq2 and Voom-Limma perform the best, with the least number of false discoveries amongst their top ranked genes up until 1250 genes; where all tests appear similar in their inability to exclude false negatives.

Further to the above, Figure 3.5 is a representation of False discovery rate (FDR). This figure indicates the number of inaccurately called false positives amongst the genes predicted to be differentially expressed at an adjusted p-value threshold of 0.05 (dashed line). A good test should see the observed FDR close or below the imposed adjusted

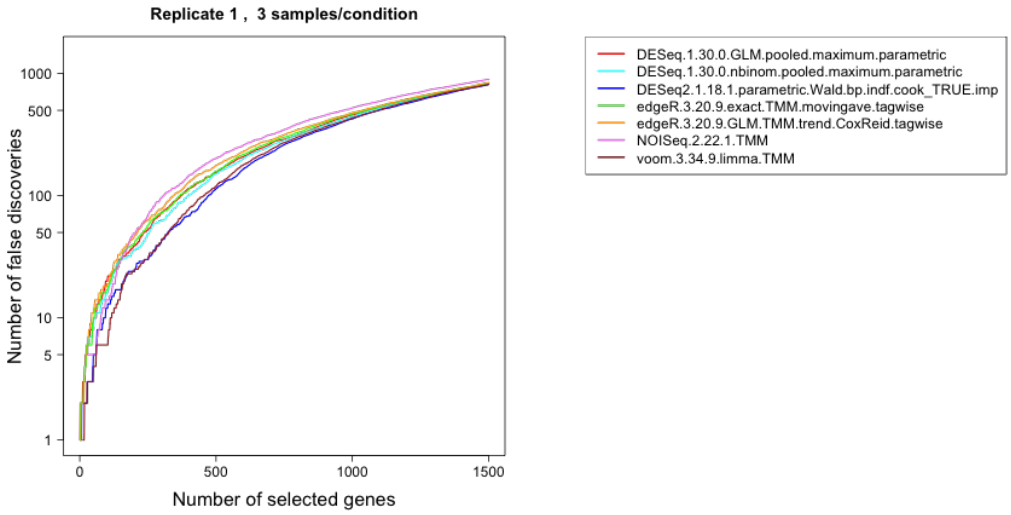


Figure 3.4: False Discovery Curves

p-value threshold. Voom-Limma appears to perform below the adjusted value threshold of 0.05, with all others greatly over this threshold.

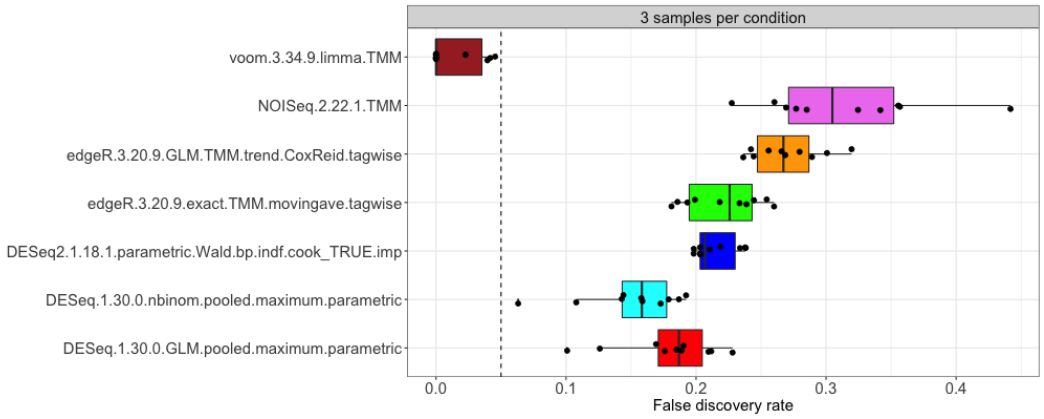


Figure 3.5: False Discovery Rate

The total number of significant genes called with the adjusted p-value threshold of 0.05 is plotted in Figure 3.6. This is compared with the originally set number of genes to be significantly different between the two conditions when synthesising the dataset (10% of total; 1250 Genes). The plot therefore is a measure of stringency alone.

In Figure 3.7, a comparison between the scores each differential expression method

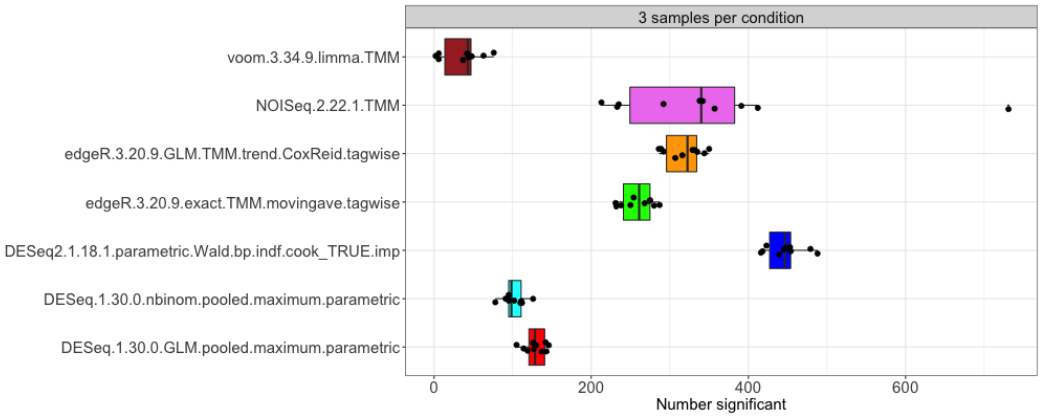


Figure 3.6: Number of Significant calls

give to each gene. The Spearman's correlation between scores returns a value between -1 and 1, with a high positive value of correlation coefficient indicating both methods similarly ranked genes. These comparisons are plotted as a heat map (index displayed). It appears that Voom-Limma shows the greatest difference between all of the other tests. Perhaps to be expected, EdgeR GLM and EdgeR exact show similar correlation.

The Matthew's Correlation Coefficient is a measure of performance summary by combining the number of observed; true positives, true negatives, false positives, and false negatives into a single metric (Figure 3.8). The coefficient is ranked from -1 to 1, with a value of 1 corresponding to a perfect classification and -1 corresponding to a completely inaccurate classification with regards to the true labels. Furthermore, on this scale a value of 0 corresponds with a random assignment of label. Here DESeq2 far outperforms the other tests in its ability to accurately reconcile all classification labels. Both Voom-Limma and the original iterations of DESeq performed the worst, with scores of around 0.23.

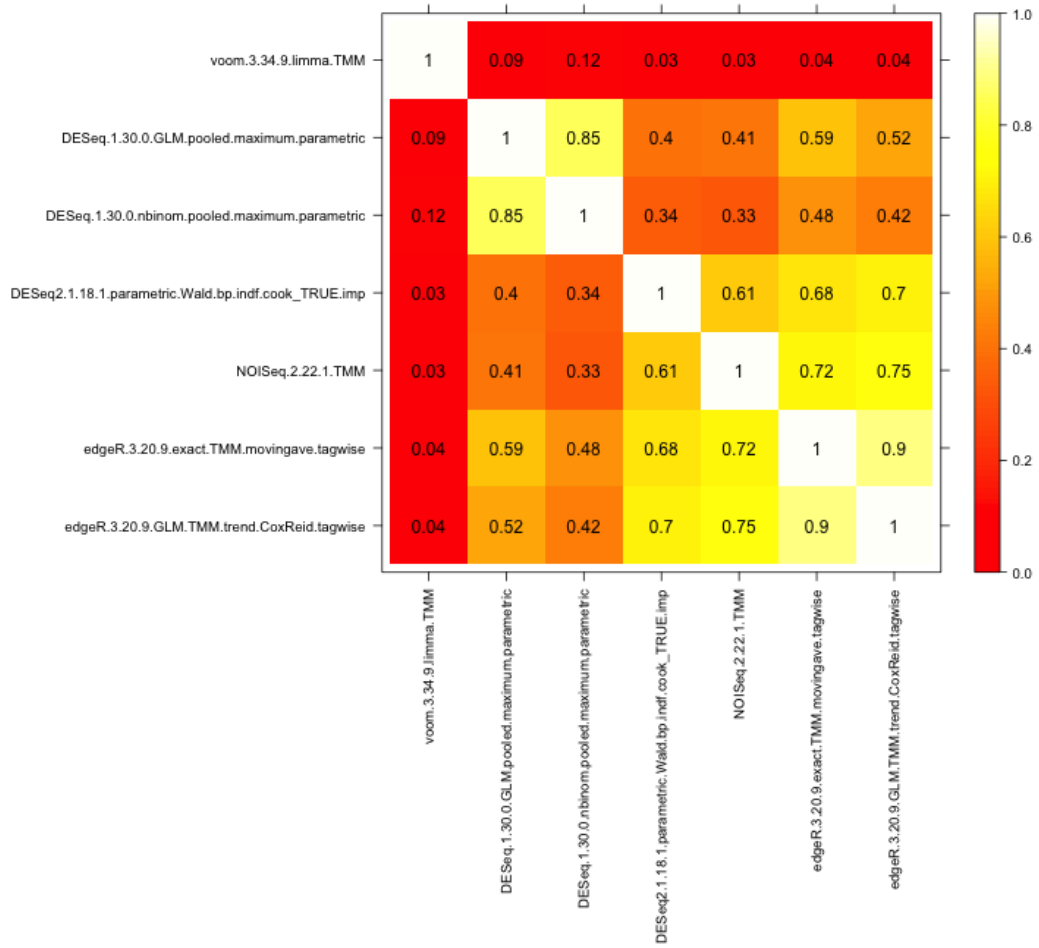


Figure 3.7: Spearman correlation between scores

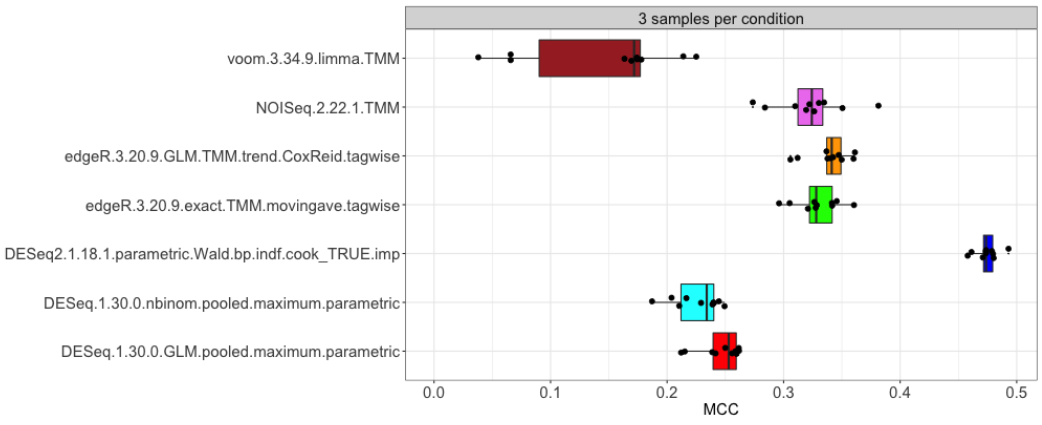


Figure 3.8: Matthew's Correlation Coefficient

Comparison of replicate number - DESeq2

Based on the above metrics, DESeq2 was selected for further optimisation to see how it best performs given a variety of replicates. A similar series of plots to the comparison of DGE methods can be seen here.

Figure 3.9 shows the AUC, based on the ROC curve outlined above. Here the AUC metric appears similarly strong across all of the number of replicates tested, with an n=6/7 showing the strongest AUC.

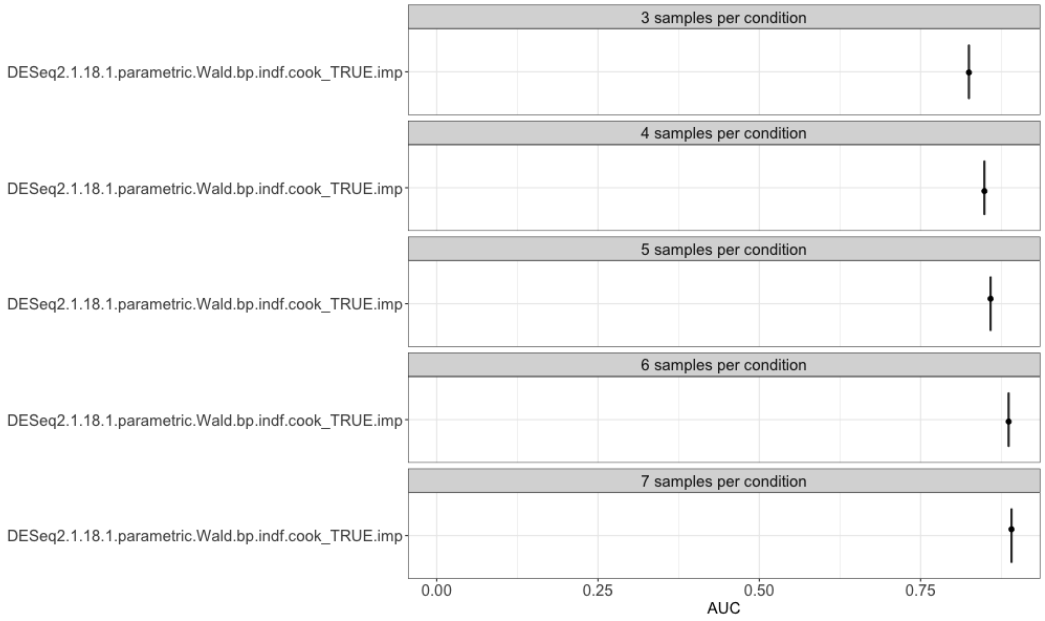


Figure 3.9: Area Under the Curve (AUC)

The false discovery rate between the samples with different replicates shows a predictable pattern, as increasing numbers of replicates lead to a significant reduction in FDR (Figure 3.10). This appears consistent, however the reduction in FDR between n=5 and n=6 is greater than the previous reductions. Furthermore, an interesting increase in FDR occurs between n=6 and n=7, showing n=6 to outperform n=7 within DESeq2.

Figure 3.11 shows a steady and consistent increase in the number of significantly

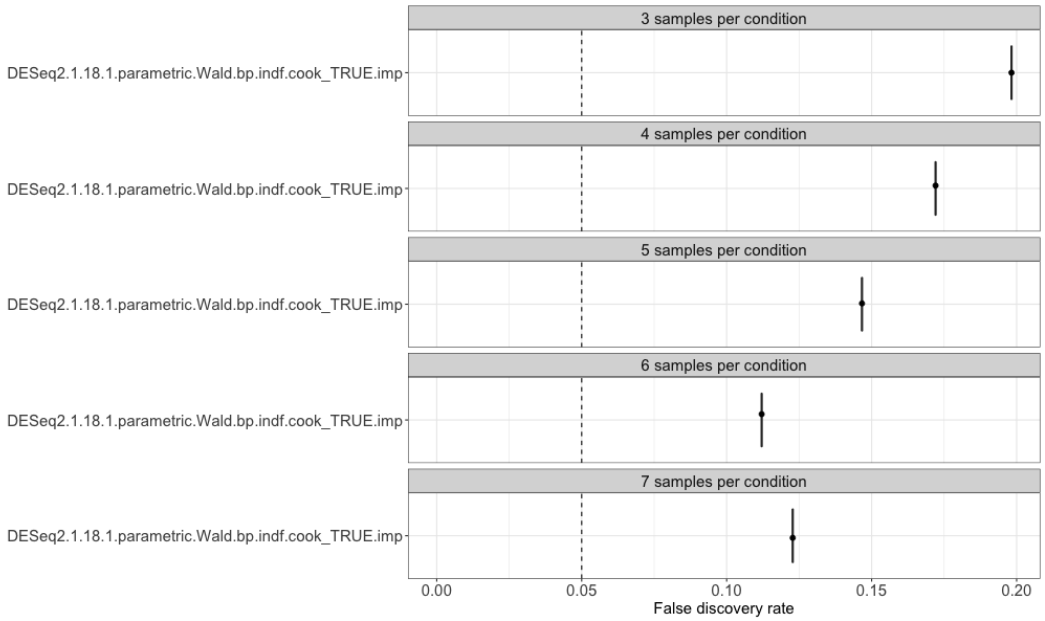


Figure 3.10: False Discovery Rate

regulated genes called as the number of replicates is increased. The *in silico* generated dataset contained a predetermined number of 1,250 significantly regulated genes. With this in mind, the closest prediction was for the n=7 dataset at approximately 800 significant calls.

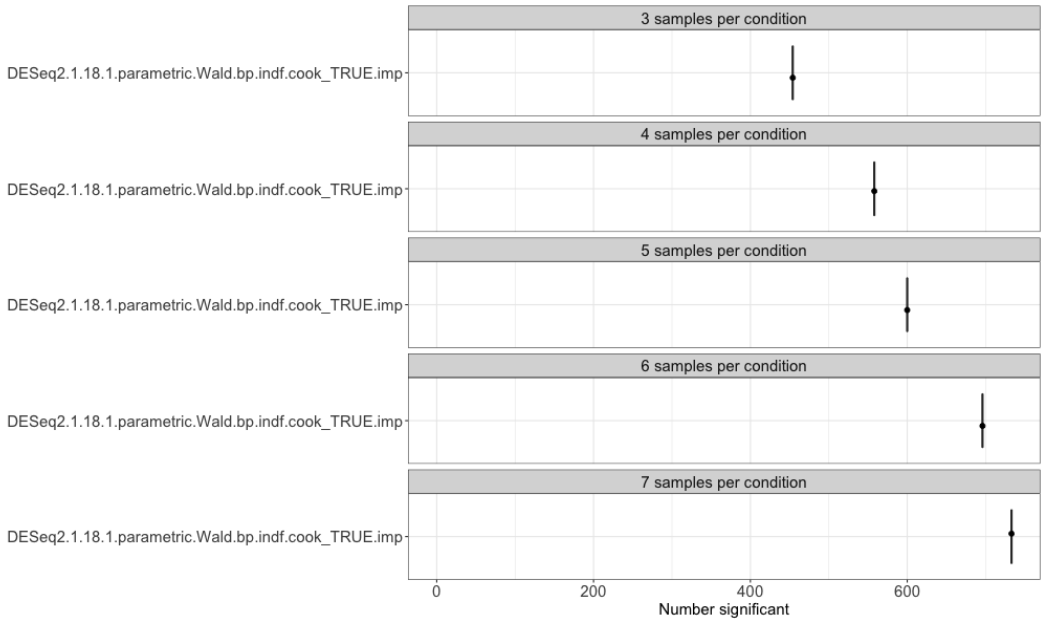


Figure 3.11: Number of Significant calls

When these data are collated by use of the Matthew's Correlation Coefficient, an interesting result appears (Figure 3.12). Contrary to expectation, an increase in replicate number does not equate to an overall better performance of the analysis. Here the n=6 dataset marginally outperforms the n=7 dataset. It should be noted that this result is based purely on a single *in silico* generated dataset, with predetermined effect size, analysed with DESeq2 only.

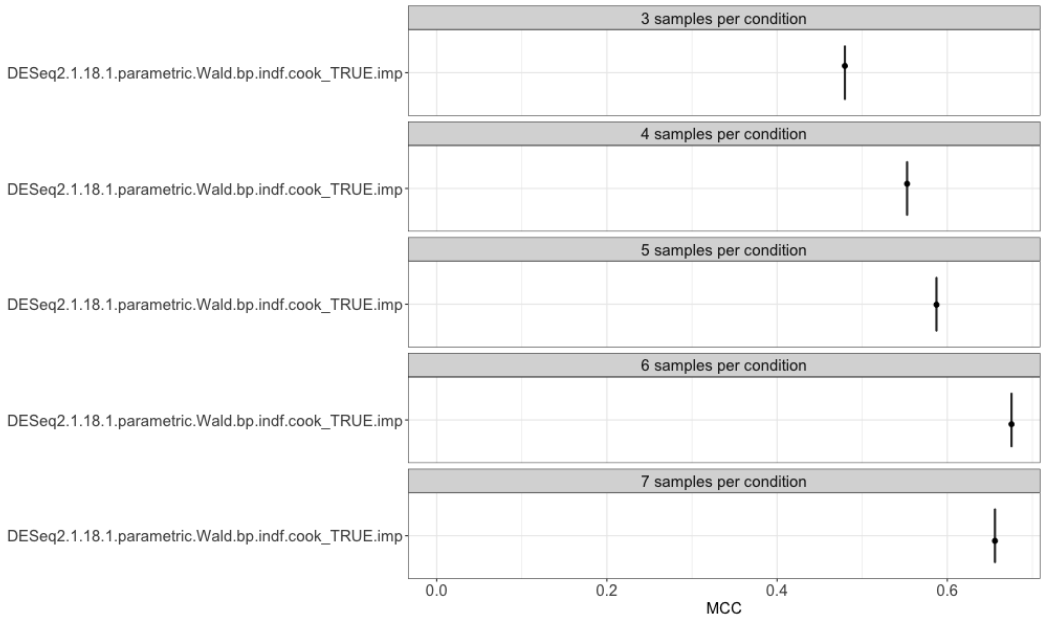


Figure 3.12: Matthew's Correlation Coefficient

Moving forward, it therefore appears that n=6 should provide a significant improvement in output for RNAseq experiments, whilst still ensuring the cost of the experiment does not become prohibitive.

3.4.2 RNAseq

RNAseq n=3 vs n=6

As with the pipeline utilised for the n=3 dataset, the initial QC output was scrutinised to ensure there were no potential errors with the raw fastq files. Here fastqc data are displayed for both the n=3 and n=6 datasets, in order to make an initial comparison between the output of the RNAseq experiments (Figure 3.13).

Ignoring the different read lengths between the two RNAseq outputs, the reads generated as part of the n=6 dataset are consistently much higher than those of the n=3 dataset (Figure 3.13). Furthermore, the levels of Phred score reduction across the read appear much less severe within the n=6 dataset.

This is further reflected within the Per Sequence Quality Scores, with the newer n=6 dataset showing a much higher number of reads with a phred score approaching 40 (Figure 3.13).

The number of N calls per base are consistently low between both datasets, however those of the n=6 dataset are only just visible as advancing over the x-axis across the read (Figure 3.13). The same cannot be said about the n=3 dataset, where the per base N content can be discerned from the x-axis. Regardless, this only ever accounts for a maximum of well below 1% of the reads, and so is unlikely to affect the alignment to the reference genome and eventual differential expression calling for genes that are abundantly present in the samples.

The per sequence GC content sees the greatest level of improvement between the datasets (Figure 3.13). The distribution for the n=6 experiment is much more consistent and even, with the fastqc issuing a warning on reads rather than a failure (as denoted by the orange colour of the distributions, rather than the red). This is due to the sum of the deviations from the normal distribution representing more that 15% of the reads. While this is not optimum, it does present a significant improvement over the n=3 experiment where the sum of deviations from the normal distribution being >30%.

The percentage of total sequences showing an over representation is markedly different between the two experiments (Figure 3.13). This is occurring despite comprehensive adapter trimming having taken place on both experimental outputs.

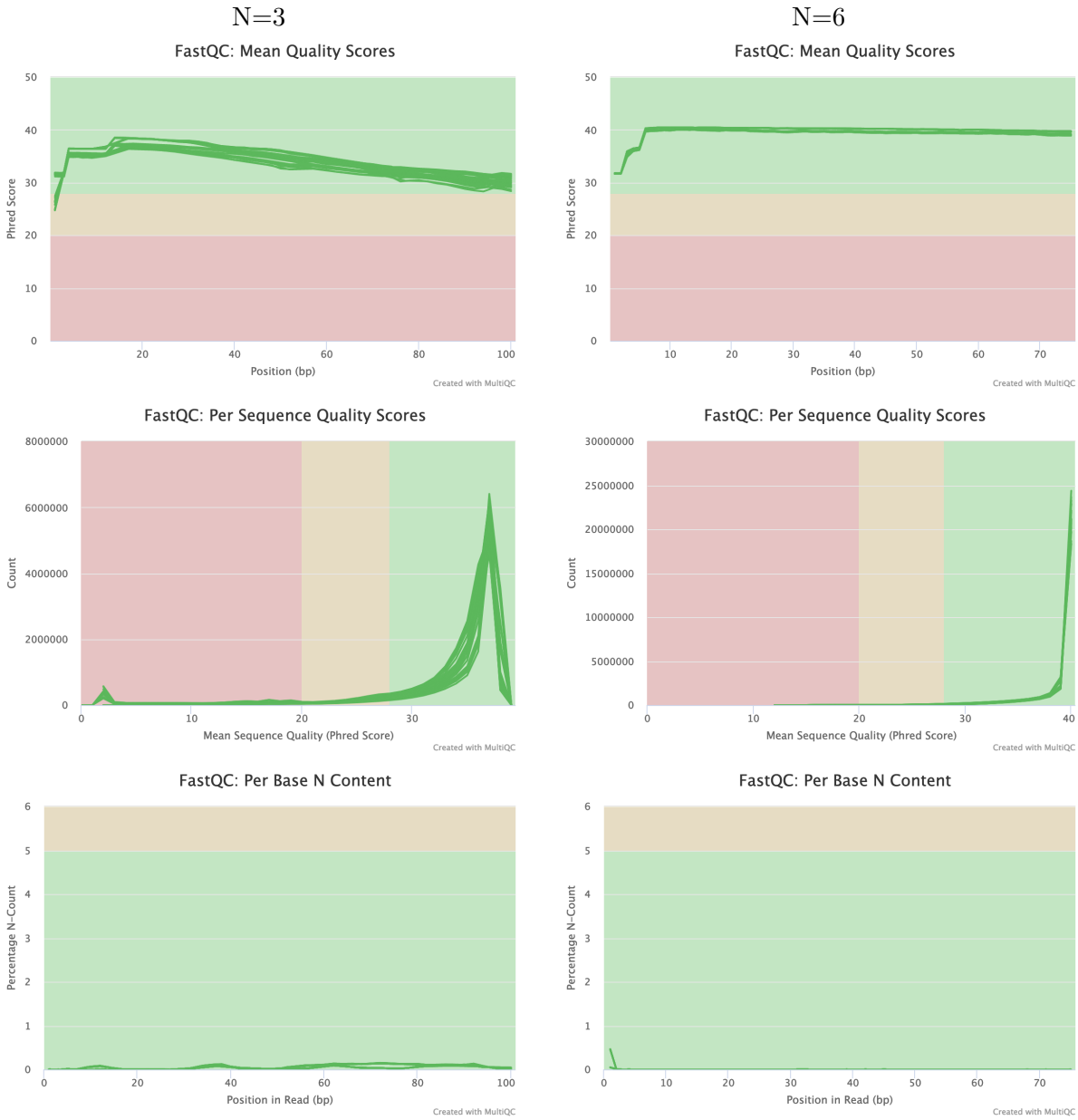


Figure 3.13: RNAseq Quality Control Metrics, as generated by the FastQC package.[67] Here Quality control metrics are displayed for both the n=3 and n=6 datasets, after adapter trimming has occured.

When the percentage of uniquely mapped reads are plotted for both Tophat and STAR, a series of distinct patterns become apparent (Figure 3.14). Most notably, the

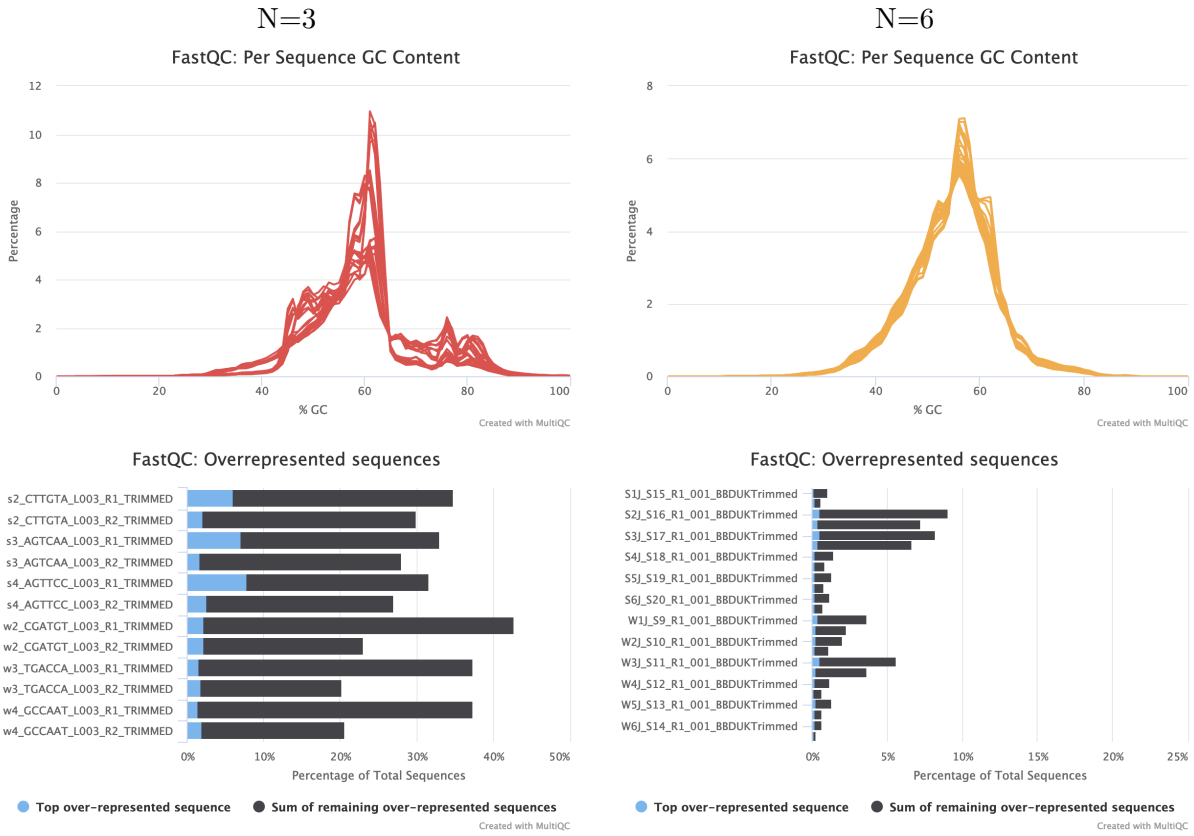


Figure 3.13: RNAseq Quality Control Metrics, as generated by the FastQC package.[67] Here Quality control metrics are displayed for both the n=3 and n=6 datasets, after adapter trimming has occured.

percentage of aligned reads for either alignment strategy are much greater for the new RNAseq experiment. When comparing alignment strategy within RNAseq experiments, a clear separation between Tophat and STAR is observed in the n=6 experiment. Within the new cohort of samples, the average number of concordantly mapped reads increases from an average of approximately 60% to just over 80% alignment between Tophat and STAR alignment strategies.

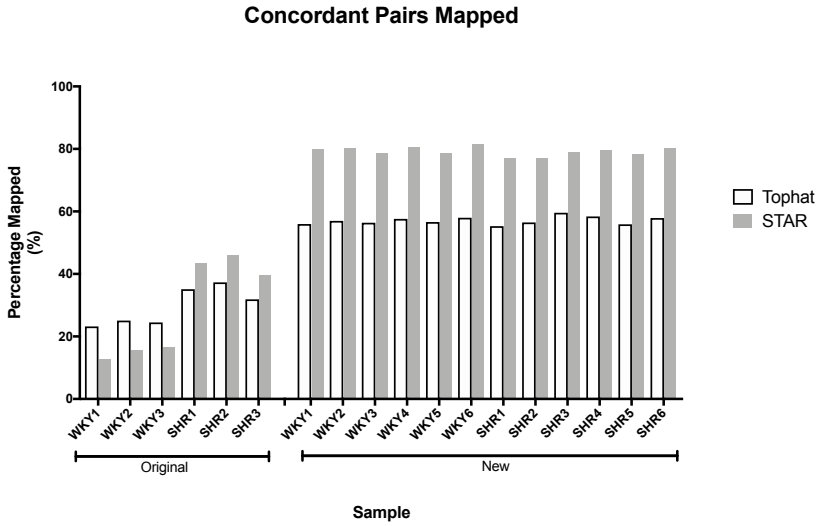


Figure 3.14: Levels of Concordantly mapped reads of both n=3 and n=6 run - Comparison between Tophat and STAR aligners

The PCA generated, as part of the DESeq2 package, shows the first two principal components to account for the majority of the variation observed between all four groups (Figure 3.15). The greatest level of variance observed (PC1; 82% Variance) clearly resolves between the old and new sequencing experiments. The second largest variance (PC2; 11% Variance), resolves between the WKY and SHR strains, regardless of the batch from which they originate. Furthermore, each individual group shows an incredibly tight clustering, showing the majority of the variance is accounted for by either strain or batch.

Figure 3.16 shows a scatter plot of DESeq2 predicted Log_2 fold changes of all genes with a multiple test corrected P-Value <0.05. A linear regression analysis shows

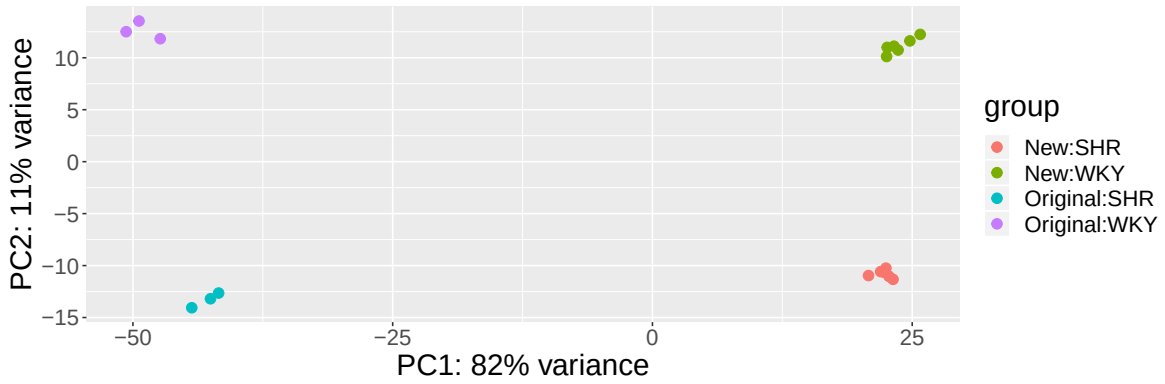


Figure 3.15: principal component analysis (PCA) Plot featuring Previous n=3 RNAseq and current n=6 RNAseq experiments. Sample colours are grouped by experiment and strain.

the correlation between the Log_2 Fold changes of both datasets with the corresponding; slope coefficient, R^2, and 95% Confidence interval (CI) values displayed. A high level of concordance appears between the values of each dataset, with the n=3 dataset showing a slightly elevated Log_2 fold change per gene as compared with the n=6 dataset.

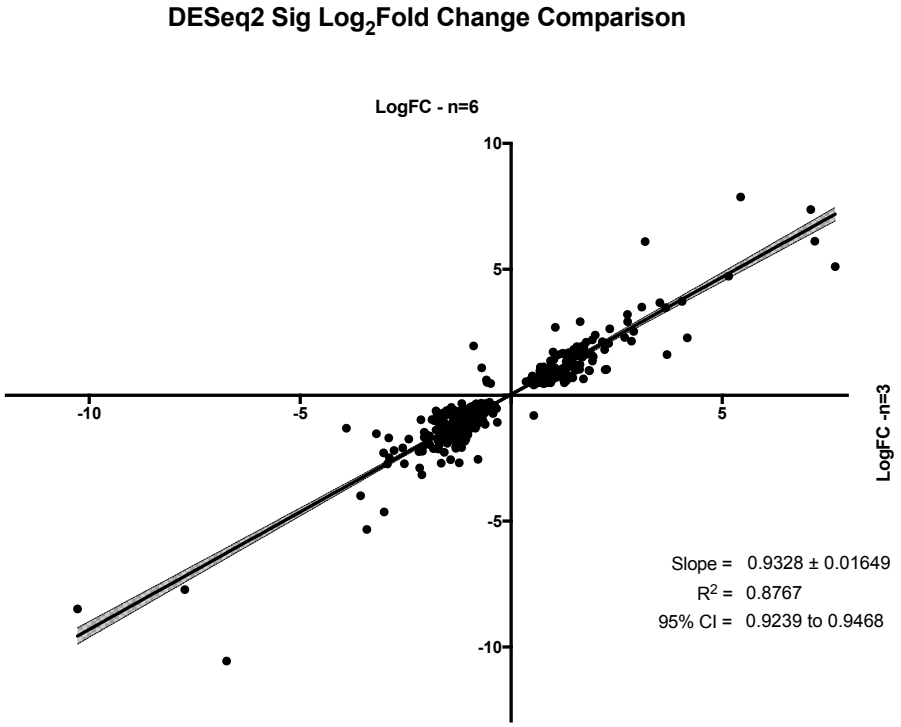


Figure 3.16: Scatter Graph representation of DESeq2 significant (P-Value<0.05) Log_2 Fold Changes across n=3 and n=6 RNAseq experiments

The total output for the new n=6 DGE analysis is displayed in Table B.2 of the appendix. Table 3.2 represents a curated version, displaying only the original candidate genes along side their associated qPCR validation values.

Figure 3.17 shows an UpsetR generated plot of DESeq2 significant (P-Value <0.05) genes for both the n=3 and n=6 datasets. Each dataset is split into positively (i.e. upregulated in the SHR as compared with the WKY) and negatively (i.e. downregulated in the SHR) regulated genes. These total set sizes are displayed on the bottom left x-axis. Intersections of conserved genes are then displayed by use of individual nodes along the main x-axis, with the number of genes common to the highlighted nodes

	DESeq2					qPCR	
	GeneName	**FC**	**P-Value (Adjusted)**	**WKY.Avg**	**SHR.Avg**	**FC**	**P-Value**
ENSRNOG00000013538	Capza1	0.98	0.97	816.72	804.06	1.1	0.60
ENSRNOG00000019050	Ifit1	165.65	7.17E-148	73.56	12195.78	63.51	<0.0001
ENSRNOG00000025108	Ankrd35	69.48	4.27E-06	4.75	330.46	3.68	0.0005
ENSRNOG00000001242	Gstt3	34.49	3.28E-94	25.46	883.96	28.47	<0.0001
ENSRNOG00000008364	Cat	1.13	0.44	31846.01	36123.92	1.6	0.09
ENSRNOG00000051682	Zcchc9	4.84	1.70E-17	55.94	269.35	6.67	<0.0001
ENSRNOG00000053450	Myadm	0.51	1.32E-12	999.85	508.76	1.35	0.04
ENSRNOG00000003977	Dusp1	0.33	9.36E-15	945.37	315.07	0.34	0.0051
ENSRNOG00000004206	Glrx5	0.40	4.06E-37	51025.59	20203.56	1.01	0.96
ENSRNOG00000021106	Gramd1a	1.32	0.04	445.87	590.12	1.58	0.003
ENSRNOG00000020938	Ppp1r15a	1.41	5.60E-05	5002.49	7034.93	1.09	0.64
ENSRNOG00000013631	Slc31a2	1.57	0.77	16.75	26.43	0.96	0.44

Table 3.2: Comparison of RNAseq Output with qPCR Validation on Candidate Biomarkers. For each candidate gene, Ensembl Gene IDs are shown, alongside the DESeq2 raw output for DGE analysis. The full table has been amended to include; Fold Change (FC), P-Value (Benjamini-Hochberg Multiple test corrected), and mean counts for each group. Also displayed are the FC and associated P-Values for the qPCR analysis.

displayed along the y-axis (as denoted by the "Intersection Size"). Highlighted in orange and blue are the shared nodes for downregulated and upregulated in both n=3 and n=6 datasets respectively. The greatest level of overlap in DESeq2 significant genes, is within the downregulated genes (321 conserved genes) compared with the upregulated genes (125 conserved genes).

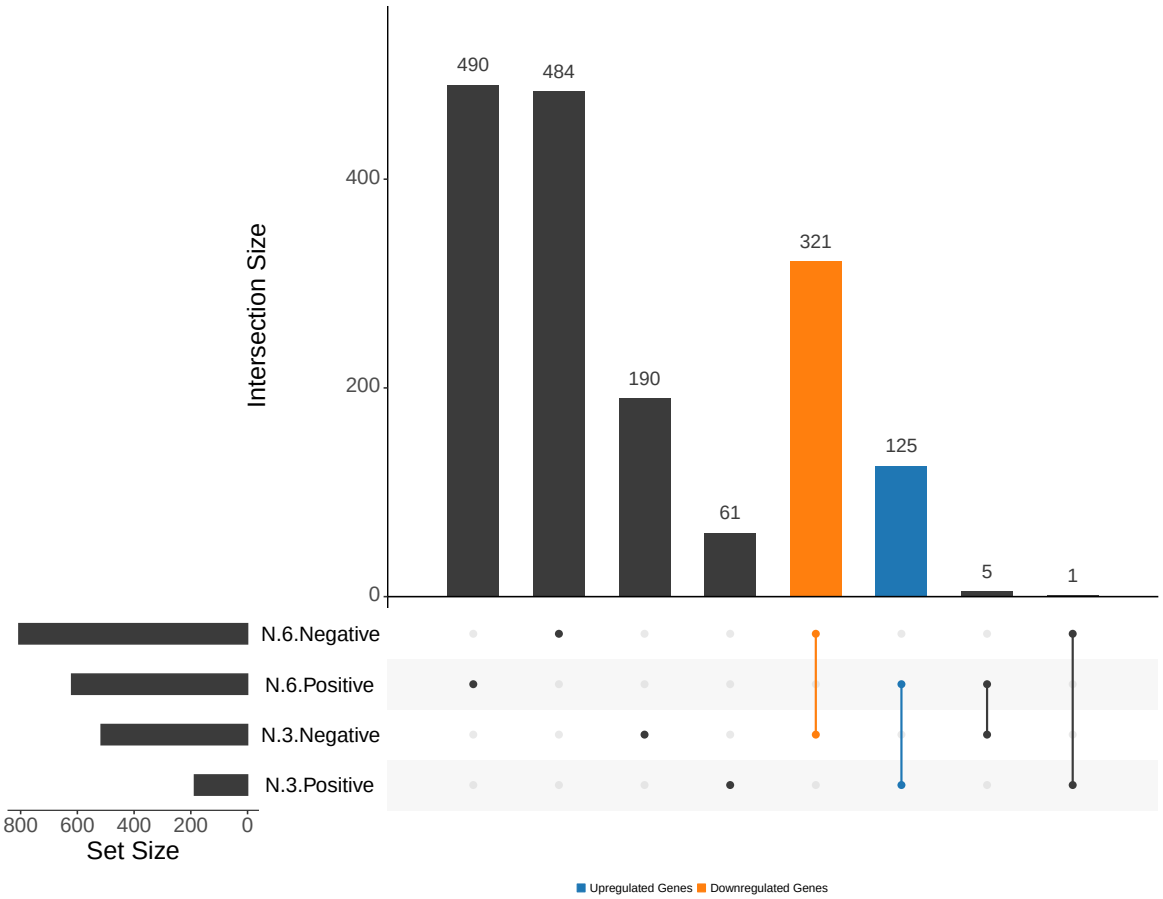


Figure 3.17: UpsetR Generated comparison of DESeq2 significant genes (P-Value<0.05) between n=3 and n=6 RNAseq experiments; separated by direction of change. Total set sizes are displayed in the bottom left x-axis, and sets being compared are designated by highlighted nodes across the main x-axis. The number of genes individual to a set, or conserved across multiple are denoted by the nodes highlighted and the corresponding "Intersection Size" on the y-axis.

3.4.3 Microarray

Quality Control

The R-generated pseudo-images showed a consistent patterning of intensities across all arrays, except for WKY5 which showed a circular artefact of increased intensity (Figure 3.18). The uniform nature of the artefact suggested it was an error with the manufacture of the chip itself. For this reason, this microarray sample was repeated, with the repeated array showing no such artefact.

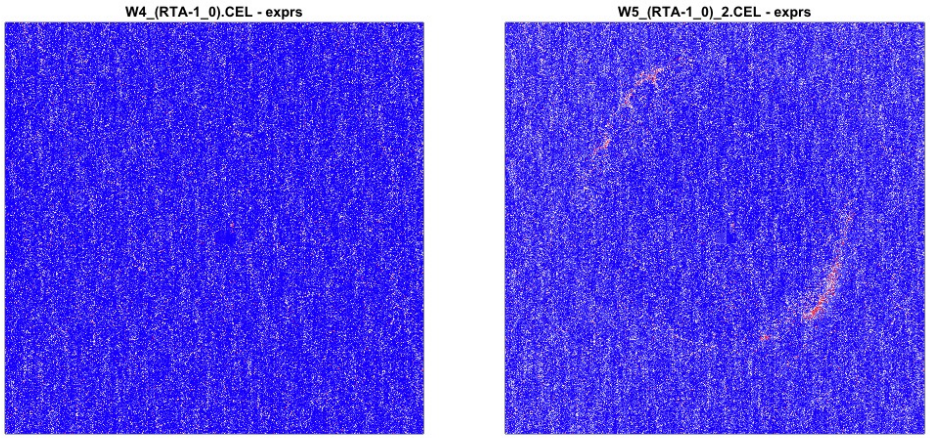


Figure 3.18: Quality control generated Pseudo-images of Microarray Gene Chips of sample WKY4 (left) and WKY5 (right). Here a clear artefact can be seen in the WKY5 sample in the shape of a ring around the chip. The artefact appears to be a defect within the gene chip, and one that was possibly the result of a loading or manufacture error. This sample was subsequently re-ran before inclusion in the experiment.

Once the array for WKY5 was rerun, a histogram representation of Log intensities across each array was produced (Figure 3.19). Here each sample shows a consistent distribution. This was therefore not indicative of any sample requiring an over normalisation or to be omitted all together.

The boxplots of the $log_2(\text{PM}ij)$ intensities and the RMA normalised $log_2(\text{PM}ij)$ intensities reveal an uneven distribution between each of the samples (Figure 3.20). This is most notable for sample SJ1 and WJ5, whose interquartile range appear trun-

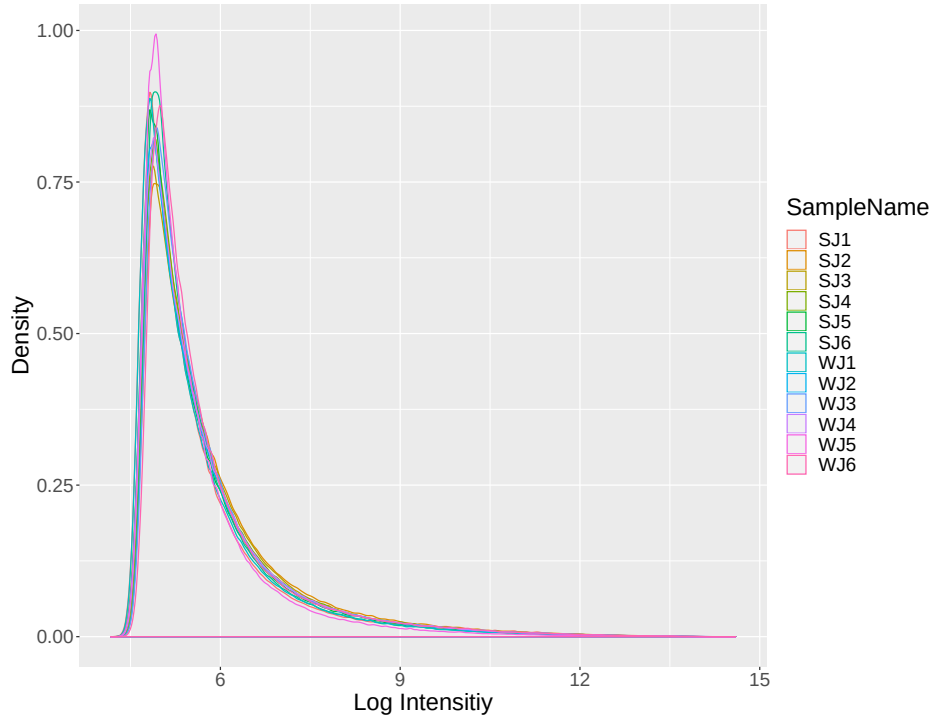


Figure 3.19: Histogram showing the Log intensities across the chip for all arrays. Here a uniform distribution of intensities across can be seen across all chips. Therefore, this QC metric does not highlight any causes for concern at this stage of the QC pipeline.

cated compared with the other samples. These intersample differences in $log_2(\text{PM}ij)$ intensities are no longer perceivable after RMA normalisation. The global distributions therefore all appear comparable.

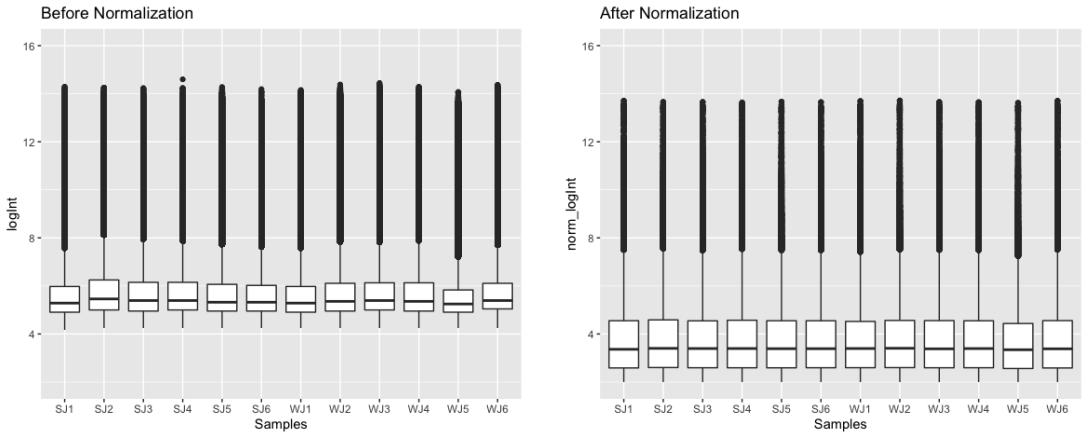


Figure 3.20: Boxplots of $log_2(\text{PM}ij)$ intensities for both pre- and post RMA normalisation.

A PCA was conducted to better assess for any overarching dimensions of noise apparent in the data, and to see whether or not they align with the experimental design (Figure 3.21). The PCA of the raw intensities shows the first principal component (i.e. the greatest dimension of variability; x-axis), to account for 67.4% of the total variation. Across this axis, the individual samples are resolved, leading to the conclusion that the greatest source of variation in the raw intensities is due to biological variability. This pattern can be observed regardless of the strain to which the samples belong. The second principal component (y-axis) accounts for 8.65% of the variation, and shows a clear resolution between the two strains across this axis.

This pattern is relatively conserved after RMA normalisation, however there is no longer a clear separation between the strains across the second principal component (y-axis; 4.83%). This dimension of variability has been normalised against, whereas the first principal component now accounts for a greater percentage of the global noise (x-axis; 74.48%). Across this axis, samples are separating based on biological or technical noise.

Differential Expression Analysis

From Thermofisher's Transcriptome Analysis Console (TAC, ThermoFisher Scientific) generated DGE table, a total of 68,011 transcript elements were detected. Of this total, 832 reached a multiple test corrected (FDR) P-Value<0.05. This total list of transcript elements comprised of 24,753 transcripts designated as "Coding", of which 442 reached a FDR P-Value<0.05 (Table B.3) of the appendix. A curated form of the final DGE output from Thermofisher's transcriptome analysis console (TAC) software for candidate genes is displayed in Table 3.3.

In Figure 3.22, the entire detected transcriptome is displayed in scatter and volcano plots. Each punctum denotes an individual transcript, with puncta highlighted in red as being significantly upregulated, and those highlighted in green significantly downreg-

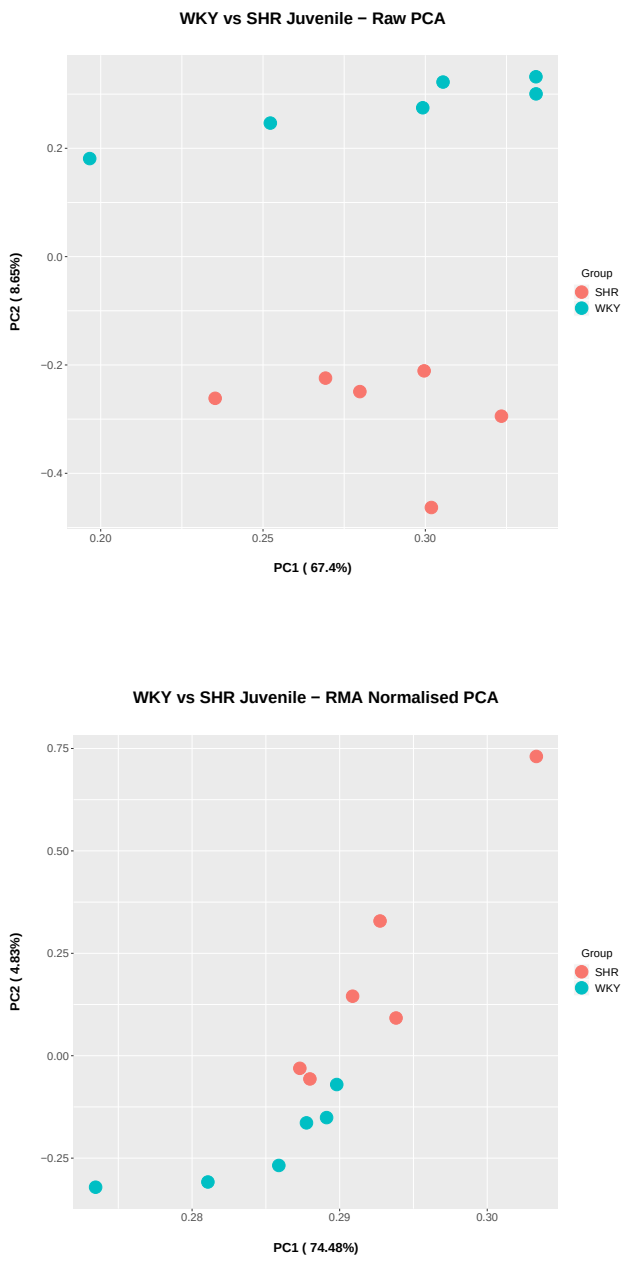


Figure 3.21: Principal Component Analysis for both pre- and post RMA normalisation. The greatest source of noise (PC1) appears to resolve the biological replicates, regardless of strain (Raw, 67.4%; RMA Normalised, 74.48%). The second principal component (PC2) separates the strains. The raw PCA shows a robust separation across the strain, which appears to be normalised for in the RMA normalised PCA (Raw, 8.65%; RMA Normalised, 4.83%).

	Microarray							qPCR	
Gene	**WKY.Avg (log2)**	**WKY SD**	**SHR.Avg (log2)**	**SHR SD**	**FC**	**P-Value**	**FDR P-Value**	**FC**	**P-Value**
Capza1	9.34	0.08	9.36	0.30	1.01	0.97	1.00	1.1	0.60
Ifit1	5.86	0.58	19.93	0.00	17179.20	2.31E-14	0.00	63.51	<0.0001
Ankrd35	4.11	0.06	4.61	0.09	1.41	6.24E-07	0.00	3.68	0.0005
Gstt3	5.27	0.11	6.69	0.14	2.67	6.18E-09	0.00	28.47	<0.0001
Cat	19.93	0.00	19.93	0.00	1.00	1.00	1.00	1.6	0.09
Zcchc9	4.84	0.13	6.77	0.17	3.80	9.67E-10	0.00	6.67	<0.0001
Myadm	7.73	0.24	6.96	0.23	0.59	3.00E-04	0.03	1.35	0.04
Dusp1	8.02	0.57	6.39	0.63	0.32	3.00E-04	0.03	0.34	0.0051
Glrx5	6.27	0.08	6.15	0.06	0.92	3.60E-03	0.17	1.01	0.96
Gramd1a	4.93	0.00	4.93	0.06	1.00	0.44	0.84	1.58	0.003
Ppp1r15a	7.60	0.07	7.85	0.20	1.20	0.01	0.27	1.09	0.64
Slc31a2	4.10	0.00	4.08	0.05	0.99	0.44	0.84	0.96	0.44

Table 3.3: Comparison of Microarray Output with qPCR Validation on Candidate Biomarkers. For each of the candidate genes, raw microarray output is shown from the transcriptome analysis console (TAC) software (Thermofisher®). Output has been curated to only show pertinent columns, including; log_2 Average signal per group, accompanying standard deviation (SD), fold change (FC), P-Value, and FDR corrected P-Value. Also displayed are the FC and P-Values from qPCR analysis.

ulated. Also highlighted here in the red circle, is the Ifit1 transcript; one of the most significantly upregulated genes in the n=3 RNAseq dataset that was validated through qPCR analysis. The scatter plot shows a greater level of genes to be upregulated in the SHR than downregulated. The volcano plot illustrates the greater level of these upregulated transcripts in reaching P-Value significance.

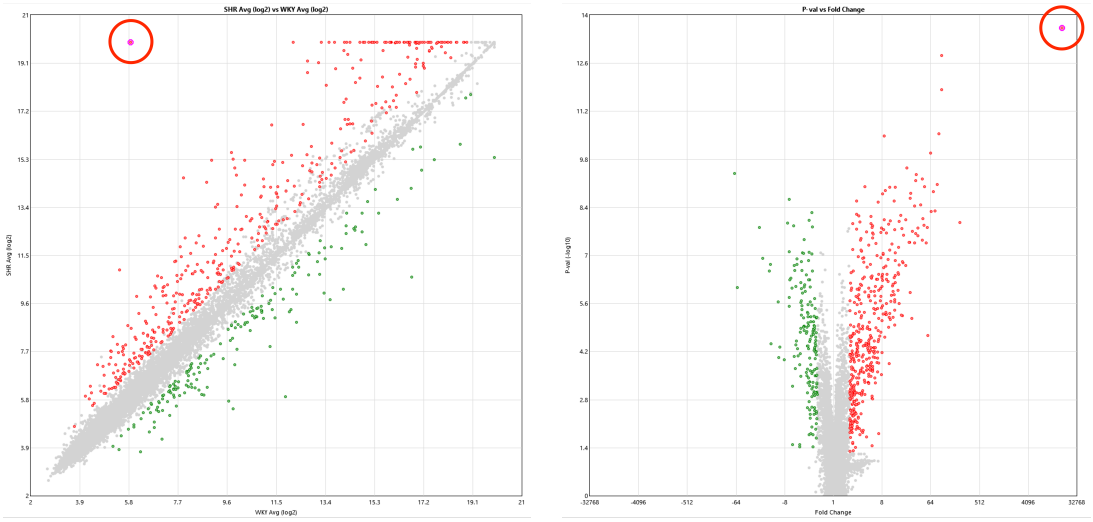


Figure 3.22: Scatter and Volcano plots of Microarray results

3.5 Discussion

3.5.1 *In Silico* Optimisation

Comparison of Methods

In an attempt to build upon the previously obtained n=3 RNAseq dataset, and mitigate the high FDR observed, an *In Silico* generated dataset was created. Using the n=3 dataset as a foundation, the parameters of the *in silico* dataset were set in order to best emulate the results of the biologically generated dataset. Unlike a biologically derived RNAseq output, the precise expression values of each gene are predetermined, allowing

for a precise correlation to be drawn between the actual and predicted DGE.

With this in mind, CompcodeR was able to produce a series of performance benchmarks for each of the statistical DGE tools. With 3 replicates, and averaged over 10 randomly generated datasets (whilst still maintaining the same parameters outlined in Table 3.1), the majority of the tests appeared consistent in their overall performance. The main benchmarks that saw a separation in the tool's performance were; the number of significant calls, the false discovery rate, and the Matthew's Correlation Coefficient.

The FDR not only varied greatly between the various DGE tools tested, but within the replicated datasets assessed. Some of the tools showed a great spread in the FDR report, such as NOISeq, and DESeq. By far the best performing tool was limma-voom with TMM normalisation. However, this was due to the hugely stringent nature of its analysis as it predicted the lowest number significantly regulated genes by a large margin.

There was a great disparity in the number of genes predicted to be significantly regulated by the various DGE tools. Interestingly, DESeq2 was markedly more stringent then EdgeR in the n=3 (Section 2.4.1; DESeq2 P-Value<0.05, 691 transcripts; EdgeR P-Value<0.05, 1,821 transcripts). However, here DESeq2 consistently predicted the greatest number of significantly regulated genes, with this number being closest to the true number of regulated genes.

In an attempt to curate these findings, and integrate them into an overall performance benchmark, a MCC was conducted. The MCC takes into account both the true and false positives and negatives, and provides a metric that is generally regarded as balanced (Figure 3.23). The produced coefficient ranges from -1 to 1, with a score of 1 denoting a completely correct binary classifier. Here DESeq2 performs markedly better, with a series of MCC scores between 0.45 and 0.5.

It should be noted that each iteration of a synthetic dataset will subtly change based on the random algorithms employed in the *generateSyntheticdata* command. This will therefore subtly influence the various comparisons drawn above. Whilst the attempt was

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{[(TP + FP) \times (FN + TN) \times (FP + TN) \times (TP + FN)]^{\frac{1}{2}}}$$

Where; TP = True Positives
TN = True Negatives
FP = False Positives
FN = False Negatives

Figure 3.23: Equation for calculating the Matthew's correlation coefficient (MCC). The MCC serves to produce a metric that integrates all of the logical outcomes from the RNAseq predictions.

made to mitigate this by running several synthetic datasets through this comparative software, each run in trialing this software has been sufficient to reliably rank each test on any given metric such that its position remains uninfluenced. For that reason, a single generated dataset can suffice in allowing the assessment of these tests as there is less concern with the absolute values they gain, rather how they perform against each other.

What is clear from these comparisons, is just how varied each strategy is in its strengths and weaknesses. No one test appears to outperform the rest across all benchmarks. However, it does appear that DESeq2 best resolves the relationship between false and true calls towards significantly regulated genes. For this reason, further optimisation made use of the DESeq2 DGE prediction software.

Comparison of replicate number - DESeq2

Based on the subtle performance differences across the DGE statistical approaches, coupled with the literature, DESeq2 was selected to further optimise the number of replicates required.

From the CompcodeR datasets, increasing the number of replicates resulted in a consistently increasing number of significantly regulated gene predictions. This follows

an expected pattern as an increase in replicates gives rise to a more accurate estimation of the sample distribution. This in turn leads to a greater statistical sensitivity in detecting truly regulated genes.

The key issue with the n=3 dataset was the high FDR. Interestingly, based on the simulated data 6 replicates leads to the lowest FDR by DESeq2. Moreover, this observation was strong enough to lead to a consistent result within the MCC benchmark as n=6 outperforms the n=7 based on MCC, contrary to expectation.

For the comparison of replicate numbers, a single replicate was generated for each DESeq2 analysis. The compcodeR *generateSyntheticdata* command works by modeling gene counts from a Negative Binomial distribution, using mean and dispersion parameters obtained from previously obtained RNAseq gene expression data.[80] A random variable is generated to represent the gene counts for this model, in conjunction with the key parameters outlined in Table 3.1. It is this random variable that will lead to interexperimental noise in gene expression counts between *generateSyntheticdata* generated data, much in the same way biological noise affects expression distributions. It is likely this noise is the cause for any discrepancies observed within the comparisons, including the counterintuitive increase in FDR in the n=7 *In Silico* dataset. This could be mitigated by replicating the replicates, and generating a larger series of datasets to be cycled through for any given replicate number to be tested.

Furthermore, it should be noted that there may be gains in performance between the different DGE statistical approaches, as the number of replicates is increased. Here only DESeq2 was benchmarked with different numbers of replicates. For this reason, a more comprehensive assessment could be ran to iteratively compare all of the DGE approaches available, with several permutations of replicates.

It would be possible to run this *In Silico*, however this has previously been performed with experimentally generated RNAseq data, as outlined in section 2.5.3. Here the focus was given to an *In Silico* generated dataset, with selected parameters to best mirror the results observed in the original n=3 dataset as no other biological proxy was

available.

Furthermore, running the analysis with additional replicates greater than 7 may shed light on the optimum number of replicates to accurately call this particular number of regulated genes, however given the increasing cost of additional replicates, this trade off may not be required for the generation of a candidate list of biomarkers.

For these reasons, and the findings of the high throughput studies mentioned in section 2.5.3, 6 replicates per group were selected for use in a repeat of the WKY and SHR blood transcriptome comparison.

3.5.2 RNAseq

n=3 vs n=6 Datasets

The fastqc output for both the n=3 and n=6 RNAseq experiments presented a series of metrics by which a comparison could be drawn between the raw fastq files of each. It is the first step after sequencing, and already shows a series of interesting differences between the experiments.

The first, and arguably most important metric of raw fastq files, was the quality of base calling from the sequencer itself. In other words, the accuracy of each base call as measured by the Phred score. Comparing across the different RNAseq experiments reveals an interesting difference between the consistency of average Phred scores across each read. The n=3 reads show a steady fall in Phred score across the read, whereas those from the n=6 experiment are consistently high. This observation could be resultant of two factors; the sequencing technology used, and the length of the reads generated.

The first potential cause of this difference, is the length of reads being used. The n=3 experiment used 2x100bp reads, whereas the n=6 dataset used much shorter 2x75bp reads. Due to the phenomenon of phasing, outlined in section 1.5.2 of the introduction, average read quality will always tend to decrease as the reads get longer. Therefore longer reads, such as those generated in the n=3 dataset, are much more

susceptible to this reduction in base calling accuracy. Of course the shorter read length comes with the caveat of reduced information. This is a concern when this additional information is required for a per base resolution required for focusing on single nucleotide polymorphism (SNPs). Here this loss of resolution is acceptable as the focus is given to whole transcripts whose expression is modified between the model organisms being studied.

Year on year, subtle improvements are being made to all next-generation sequencing platforms. Most notably with Illumina's sequencing by synthesis technology, improvements to the sequencing chemistry and the methods of detection are constantly resulting in increasing accuracy to base calls. Here the n=3 dataset was generated with the HiSeq2500 (Illumina, Inc. - San Diego, CA, USA), a platform that was released in 2012. The n=6 experiment was conducted on the HiSeq4000; a platform released 3 years later in 2015. This difference in time translates to a great increase in base call accuracy. One of the most notable improvements to Illumina's approach is the transition from standard flow cells to "Patterned" flow cells, as used in the HiSeq4000 (Figure 3.24). Patterned flow cells contain billions of nanowells at fixed locations across both surfaces of the flow cell. This structured organisation leads to an even spacing of sequencing clusters across the the cell, enabling a clearer detection of bases from each cluster and mitigating the previous need to map cluster sites. This results in higher Phred Q-Scores and a quicker run time.[123]

One of the main differences between the n=3 and n=6 RNAseq experiments with regards to QC output, was the greatly differing GC content per sequence. FastQC issued a complete failure for the n=3 reads due to the sum of the deviations being >30%. This was not the case for the n=6 experiment, where only a warning was issued (based on the FastQC threshold of >15%). The main reasoning behind this could be the improved approach towards good laboratory practice. This way potential bacterial contaminants with differing GC contents could be successfully avoided.

Another large discrepancy between the experiments was the difference in overrepre-

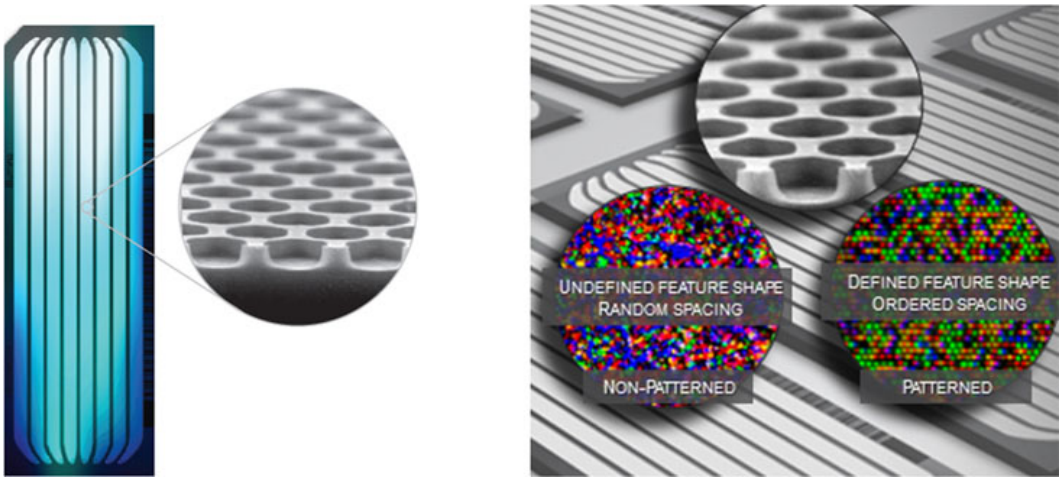


Figure 3.24: Illustration of Illumina's Patterned Flow Cell Technology. Each flow cell contained billions of nanowells, each forcing clustering to follow an ordered spacing. This translates to improved confidence in base calling due to the lack of overlapping signals being produced.

sented sequences. This appears to be resultant of the method in which the libraries were prepared. The n=3 experiment made use of ribosomal RNA depletion, whereas the n=6 used poly(A) selection. This seemingly subtle difference in obtaining mRNA transcripts translates to a significant difference in the output of libraries for sequencing. Following a nBLAST of the sequences flagged as being overrepresented, the majority of those observed in the n=3 dataset were found to be rRNA or mtRNA in nature. This is due to the method of rRNA depletion's inability to efficiently remove all rRNA from the sample before sequencing. As with any molecular protocol there will never be a 100% recovery or depletion of a molecular target, meaning several rRNA transcripts will remain in the library for sequencing.

This problem is mitigated by the Poly(A) selection method, as even a suboptimal recovery of the total poly-adenylated mRNA population will still be relatively devoid of rRNA. This way, the finite amount of RNAseq reads per flow cell can be utilised for identifying protein-coding genes. Of course this prevents the discovery of novel non-coding transcripts that are increasingly found to have a profound regulatory role in biology.[124] However, here the focus is on protein coding transcripts, and so the Poly(A)

selection method presents the best option for moving forward.

The decision was made to optimise the step of read alignment, in order to tease out the greatest level of information regarding gene expression from the data. Here the previous pipeline for the n=3 dataset made use of the Tophat read alignment algorithm. The newer spliced transcripts alignment to a reference (STAR) algorithm was selected to compare Tophat against. When assessing the alignment rates between Tophat and STAR, the Original and New experiments must be taken in isolation as different experimental approaches, including the differing library preparations, will translate to differences in reads produced and ultimately how they are aligned. Taking the new n=6 dataset in isolation, a clear separation in concordant alignment rates is observed between Tophat and STAR. The writers of the STAR algorithm claim a higher level of sensitivity, while maintaining a low false-positive rate, due to the way in which STAR processes reads.[70]

As part of its approach, STAR can be split into two distinct phases of alignment; seed searching, and clustering/stitching/scoring. The initial phase involves taking each read and searching for the longest sequence that it can match in one or more locations on the reference genome. The longest perfect match becomes known as the maximum mappable prefix (MMP). The aspects of the reads that can be mapped separately are referred to as "seeds", so the first maximum mappable prefix (MMP) becomes *seed1*. STAR then searches for the remainder of the read to find the next MMP which it terms *seed2*. Each MMP search is conducted in both forward and reverse directions, facilitating a better ability to find anchors for reads the contain errors towards their ends, improving mapping sensitivity for higher sequencing error rate conditions.

The second part of STAR's approach takes all of the generated seeds and stitches them together by grouping all seeds that are not multimapping, terming them 'anchor' seeds. These are used for producing the final alignments. The algorithm importantly clusters seeds from the mates of paired-end RNAseq reads, using each paired-end read as a single sequence. This allows for a possible genomic gap, or an overlap between the

inner ends of the paired-end mates. It is this principled usage of paired-end data that leads to a further increased sensitivity towards read alignment.

A detailed comparison, conducted in 2017 by Costa-Silva *et al.*, took two human RNAseq datasets and ran them through a permutation of DGE pipelines. As part of this study, six methods of mapping reads were used, including Tophat and STAR. Based on a series of benchmarks, including total number of significant calls and concordance with qRT-PCR data, the group concluded that the method by which the reads were initially aligned did little to impact on the final analysis.[125]

However, it seems unlikely that such a large difference between the concordantly mapped pairs, observed here between Tophat and STAR, would not translate to important differences in eventual DGE results. As the paper does not declare the mapping rates for each of the sequence aligners used, it is not possible to see whether any discrepancies in alignment rate are apparent or indeed whether they are comparable to those observed here. Due to the higher levels of concordantly aligned reads STAR presents the best way in which to proceed for an optimised RNAseq pipeline.

The differences observed within the QC metrics previously discussed are all methods of assessing the suitability of the raw reads for eventual differential expression analysis. Therefore, regardless of whether they pass or fail these metrics, robustly regulated genes may still ultimately surface to be predicted as significantly regulated.

A good way in which to further probe the data, and assess for levels of noise arising from differing expression values, is to conduct a principal component analysis (PCA). Here an interesting series of observations can be made. The first observation, is that of the first principal component which appears to best resolve between the n=3 and n=6 datasets. This means that the global noise, irrespective of source, is due to batch effects observed between the two experiments. While from the same biological strains at comparable ages, these samples have been subjected to completely different library preparations and processing pipelines. Therefore, it is not surprising that batch effects represent a greater source of variability than the differences between the strains

themselves.

3.5.3 MicroArray

For this experiment, RMA was used to normalise the data at the beginning of the pipeline. It is possible to use other methods of microarray normalisation, such as the Gene-Chip RMA (GCRMA). GCRMA is an expansion on RMA, with the only difference being the internal correcting for non-specific binding to the probes, contrary to the RMA approach which does not factor in the issue of non-specific binding. It does this by using the probe sequence information to estimate a ratio of probe affinity to non-specific binding. During a series of pilot experiments, the developers of the algorithm determined the contributions of each nucleotide on each position of the probes, making use of artificial DNAs that mismatched at specific positions.[126] As there was no specific binding within this experiment, the only intensities measured were those resulting from mismatching. This enabled the team to estimate the affinities of each of the probes. GCRMA can either use this experiment as a reference, or compute new probe affinities based on the negative control probes of the specific array used.

Here the RMA approach was sufficient to produce a normalised dataset for use in differential gene expression analysis. All QC output observed was based on this approach, with RMA appearing to adequately normalise the global data expression ranges. A repeated analysis could be ran using the GCRMA approach to account for the non-specific binding, however this comparison is beyond the scope of this thesis, and the lower level of normalisation (RMA) has already demonstrated the microarray approach to be suitable for differential expression analysis in this context.

3.5.4 RNAseq vs MicroArray

Interestingly, the microarray dataset presented no false positives, and so had a FDR of 0%. While this is a large reduction in FDR from the n=6 RNAseq dataset, microarray

as a technology is still prohibitive for downstream analysis. The slightly higher FDR observed in the DESeq2 analysed n=6 RNAseq data (0.28%), is an acceptable trade-off when considering the additional benefits RNAseq offer. Through the very nature of RNAseq's sequencing by synthesis, one of the key advantages of RNAseq is its ability to differentiate the expression abundances of individual isoforms of transcripts. Moreover, if a library is sequenced with a sufficient read depth, discrepancies in reads at the base resolution can aid in single nucleotide polymorphism (SNP) discovery. Furthermore, and providing the use of a well annotated reference genome such as here with the *Rattus norvegicus*, interpretation of read counts is intuitively simple compared to the intensities given by microarray chips.[125] These additional advantages allow RNAseq technology to take this research from biomarker discovery to biomarker characterisation. For this reason, RNAseq will be used as the foundation of biomarker discovery moving forward.

It is worth noting that here 12 genes were selected for validation of the RNAseq and Microarray datasets. This number represents a modest fraction of total genes to validate such high-throughput methods as RNAseq and Microarray. In order to gain additional confidence in each dataset's FDR, and by extension their validity as a whole, additional candidates must be validated with qPCR technology. Individual transcript qPCR analysis is time consuming, as primers must be designed and validated before using to assess individual transcript expression across replicates.

A commercially available qPCR plate array would allow for a much greater number of transcripts to be analysed in a high-throughput manner. Furthermore, each of the primers are previously optimised and need to meet the stringent performance standards that follow the minimum information for publication of quantitative real-time PCR experiments (MIQE) guidelines.[125] These plates do come at a much greater cost, however this approach would greatly reduce the time taken to validate high-throughput transcriptome analyses, and provide a series of expression metrics from an entirely separate technology. The completely different experimental pipeline, and the high-

throughput approach, could provide a standard workflow to be conducted side by side with RNAseq/Microarray analysis, ultimately leading to a final sequencing dataset with a much greater level of confidence. At the time of writing, Thermofisher® offer 842 TaqMan® Gene expression assays for Rat alone (86 when applying the search term of "blood"). With this specific project in mind, the TaqMan® Array Rat Hypertension 96-well plate (Array ID; RAWCWC9) would be an ideal starting point for blood specific expression assay.

3.5.5 Concluding remarks

This chapter has allowed for a much greater understanding the potential pitfalls of high-throughput data, and has indeed been successful in mitigating many of these. Through a variety of changes to the RNAseq pipeline, including; a new library preparation method, additional replicates, and a modified analysis pipeline, a much more established foundation can be lay for biomarker discovery. Here and regardless of the technology used, the Ifit1 transcript saw the most robust increase in the blood of the pre-hypertensive SHR compared with the normotensive WKY. This increase was further validated through qPCR in section 2.4.2. Due to this huge increase, focus will be given to better understanding this particular transcript's expression profile within this model of hypertension.

Chapter 4

Validation of Candidate Selection in WKY and SHR Strains

4.1 Introduction

In this chapter, experiments have been designed to build upon previous findings by validating Ifit1 expression in a new cohort of animals from an isolated breeding colony. Furthermore, by using the optimised RNAseq dataset efforts can be made to identify and validate a novel reference gene for qPCR analysis. The expression profiles observed in this new cohort can then be linked to physiological changes taking place within the hypertensive rat strain to gain a direct correlation between physiological metrics and expression data.

4.1.1 Genetically Isolated Animal Cohort

The previous identification and validation of Ifit1 was conducted in a cohort of animals bred by a specific supplier, situated within the U.K. (Envigo). However, a notable level of behavioural and genetic differences have been observed within both the WKY and SHR strains, with differences often being attributable to the vendors and breeders from

which they originate.[127] As an extension of assessing Ifit1, and its ability to discern between a normotensive and hypertensive rat, the following experiments were conducted in animals from a geographically isolated breeding population at the Multidisciplinary Center for Biological Research (UNICAMP - São Paulo, Brazil) that also display a similar health status to the U.K. population previously studied.[128] This was in order to discern whether these robust changes were common to other substrains and not simply a quirk of the substrains previously tested.

4.1.2 Spectral Analysis & Baroreceptor Sensitivity

With technological and mathematical advances, it has become possible for blood pressure readings to reveal more than simply systolic, diastolic blood pressure or heart beats per minute. Mean arterial pressure is a culmination of mainly; cardiac output and systemic vascular resistance.

In the early 1970's, Sayers *et al.* measured the power spectrum of heart rates to elucidate their frequency content. The team showed that as well as the previously observed relationship between heart rate fluctuations and respiratory cycle, there are additional periodic fluctuations in heart rate that occur at lower frequencies.[129,130] Further work has shown that low-frequency peaks in heart rate fluctuations are related to cyclic fluctuations in peripheral vasomotor tone, whereas mid-frequency peaks are related to the frequency response of the baroreceptor reflex.[131]

A decade later, Akselrod *et al.* observed a link between drugs that autonomically modulate the heart, and fast frequency oscillations in the heart rate spectrum, suggesting the technique could be useful in the quantification of heart rate control via the autonomic nervous system (ANS).[132]

In humans and rats, three distinct frequency domains have been noted; Very Low Frequency (VLF; 0.0195 - 0.25Hz), Low Frequency (LF; 0.27-0.74Hz), and High Frequency (HF; 0.75 - 5Hz).[133] Each of these components reflect a distinct aspect of

physiological control of blood pressure.

The VLF component of blood pressure contains the majority of the signal spectra, and has been found to be reflective of multiple mechanisms of blood pressure regulation such as; myogenic tone, physical activity, temperature regulation, and efferent sympathetic outflow. Furthermore, various circulating or local vasoactive factors including Nitric Oxide, and Vasopressin have been found to alter the VLF spectra observed.[134–136]

The LF oscillations of SBP, DBP, and the LF/HF ratio of HR are all recognised markers of sympathetic drive towards blood vessels and the sympatho-vagal balance within the heart, respectively.[137–139] Furthermore, the LF component can become especially predominant during cardiovascular stress, indicating a heightened sympathetic outflow and greater baroreflex response.[134]

HF components represent a more minor constituent of the total spectrum. For blood pressure, they reflect the respiratory effects on the large thoracic vessels. Within the heart rate spectrum, the HF domain constitutes the greatest power of the whole spectrum, and arises as a result of vagal-mediated regulation of heart rates due to respiration (e.g. Respiratory Sinus Arrhythmia).[133,134]

Spectral analysis therefore serves to decompose these complex constituents of BP and heart rate readings, and display them as a function of frequency (Figure 4.1). The central methodology of spectral analysis can be reduced to either non-parametric (Fast Fourier Transformation) or parametric (Auto-regressive modelling), however these mostly provide similar outcomes.[133] Here focus will be given to the LF component of blood pressure signalling, in an attempt to assess the sympathetic drive associated with biomarker expression.

The baroreceptor reflex is an important regulator of the cardiovascular system. It functions to maintain arterial pressure within physiological levels through negative feedback, in order to maintain a homeostatic set point. Through baroreceptors in the aortic arch and carotid sinuses, detections of high blood pressure trigger afferent pathways towards the NTS. These signals ultimately culminate in modulating sympathetic and

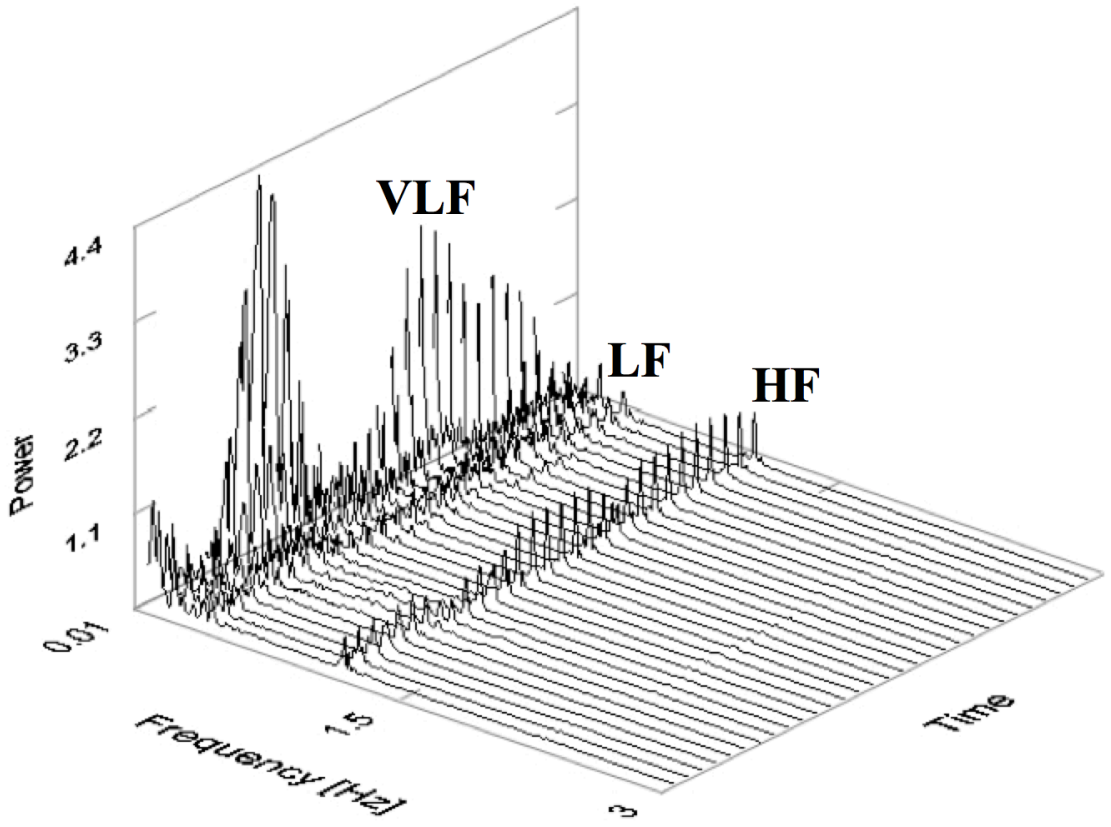


Figure 4.1: Representative plot of blood pressure spectra illustrating the three distinct frequency domains observed following a Fast Fourier Transformation of blood pressure signal; Very Low Frequency (VLF; 0.0195 - 0.25Hz), Low Frequency (LF; 0.27-0.74Hz), and High Frequency (HF; 0.75 - 5Hz).

parasympathetic outflow, resulting in continuous alterations to stroke volume, heart rate, and systemic vascular resistance.[39]

Many cardiovascular diseases have been associated with alterations towards baroreceptor sensitivity, for example; chronic heart failure, myocardial infarction, and obstructive sleep apnea.[140–142] Within the study of hypertension, baroreceptor regulation has been shown to "reset" towards an elevated blood pressure set point and to function with a reduced sensitivity towards hypertension.[143]

Here spectral analysis and baroreceptor sensitivity analysis are employed in order to resolve additional information on the cause of hypertension in the SHR. Furthermore, to use these to see whether Ifit1 can be used as a correlative measure of spectral analysis or baroreceptor derived metrics of cardiovascular disease.

4.2 Aims

To link transcript expression data at juvenile and adult ages with physiological measurements of blood pressure by;

- Establishing an experimental design to gather direct blood pressure measurements from alive animals.
- To make use of spectral analysis methods in order to tease out subtle differences in neurological control in blood pressure across the different ages.
- To link these metrics back to transcript expression data to see whether their expression reflects physiological changes taking place.
- To make use of the previously obtained RNAseq data to select a novel reference gene for further qPCR normalisation.

4.3 Materials and Methods

4.3.1 Animals

All subsequent experiments were conducted by myself. All animals used for the following procedures were sourced from the Multidisciplinary Center for Biological Research (UNICAMP - São Paulo, Brazil). Their usage adhered to the Guide for the Care and Use of Laboratory Animals from the National Academy of Sciences, published by the National Institute of Health (Copyright © 1996 by the National Academy of Sciences) and was ethically approved by the Committee of Ethics in Animal Research from Ribeirão Preto Medical School, University of São Paulo, SP, Brazil. All animals were housed individually with *ad libitum* access to food and water, and were maintained on a 12 hour light-dark cycle at a maintained temperature of 21°C.

4.3.2 Surgical Procedure

Rats were initially anesthetized with a mixture of ketamine (50mg/kg, I.P.; União Química Farmacêutica Nacional S/A, Embu-Guaçu, SP, Brazil) and Xylazine (10mg/kg, I.P.; Hertape Calier Saúde animal S/A, Juatuba, MG, Brazil), and implanted with a polyethylene catheter (PE-50 attached to PE-10, Becton Dickinson, Sparks, MD, U.S.A.) into the left femoral artery. The catheters were tunneled via the subcutaneous space and through an incision in the interscapular region of the animals' nape.

4.3.3 Direct Blood Pressure Measurement

After 48 hours recovery from the surgical procedures, the arterial lines from the animals were connected to a pressure transducer (MlT844, ADInstruments - Bella Vista, NSW, Australia) attached to a Bridge Amplifier (FE221, ADInstruments - Bella Vista, NSW, Australia). Animals were housed individually and allowed to freely roam their cages as arterial pressure was continuously sampled at a rate of 2kHz for 30 minutes in an

IBM/PC through an analogic to digital interface (Power Lab 4/40, ADInstruments - Bella Vista, NSW, Australia).

Using a three-way valve within the measurement apparatus, a bisecting section of tubing attached to a pump and pressure meter was used to make measurements at various defined pressures to calibrate the resulting mV signal range detected by the amplifier. These values were then used to calibrate the signals for each rat, enabling a conversion to be made from detected mV to an interpretable mmHg. From this, an average MAP, SBP, DBP and BPM could be drawn from stable regions of signal.

4.3.4 Spectral Analysis & Baroreceptor Sensitivity

Prior to spectral analysis, BP and HR signals were re-sampled at 20Hz and processed with a nine-point Hanning window filter and linear trend removal as detailed by Japundžic-Žigon *et al.*[144] Using a fast fourier transformation (FFT), spectra could be obtained across 15 overlapping 2048-point time series. This corresponded to 410 seconds (approximately 7 minutes) of sampling time for SBP, DBP, and HR. The power spectrum for BP ($mmHg^2$) and HR (BPM^2) was calculated for 30 FFT segments across the whole spectrum (0.0195-3Hz) and for the following frequency ranges; High Frequency (HF, 0.8-3Hz), Low Frequency (LF, 0.195-0.8Hz), and Very Low Frequency (VLF, 0.0195-0.195Hz). BP and HR spectra were then expressed in normalised units in order to assess spectral frequency distribution. The oscillation of LF signal for SBP, DBP, and LF/HF HR are established biomarkers of sympathetic regulation of vascular tone and sympathovagal balance to the heart, respectively.[138,139]

Here the Spontaneous Baroreceptor Reflex by the sequence method was employed to scan the signal for consecutively increasing/decreasing SBP samples, followed by a series of increasing/decreasing pulse interval (PI) samples that were delayed by 3 - 5 beats with respect to SBP. A threshold of sequence length was set to 4 beats.[143] Using in-house developed software (BP Complete), baroreceptor sensitivity (in ms $mmHg^{-1}$)

was asssessed as a mean linear regression coefficient (PI = BRS x SBP + Constant; where fitting of the regression model is conducted using the "sum of least squares" method). Baroreceptor set point for SBP (in mmHg) was calculated as the median value of all SBP-PI sequence points, as defined in Bajić *et al.*[145] Comparison of baroreceptor sensitivity and set point between the strains at either age was conducted as an unpaired Student's T-Test.

4.3.5 Linear Regression Modeling

All qPCR work was conducted in the University of Bristol, U.K. on blood samples shipped from the University of São Paulo and was conducted as previously outlined. Linear regression modeling was conducted on raw Ifit1 C_T levels against various physiological metrics measured here using GraphPad Prism (Version 7.0, La Jolla California USA,[146]). The software iteratively calculated the values of slope and intercept of several overlaid linear lines of best fit, in order to produce a regression model that minimises the sum of squares of the vertical distances of the points from the line.

4.3.6 New qPCR Reference Gene Selection

In order to normalise the final abundance measurements from the $2^{-\Delta\Delta C_T}$ technique, as outlined in section 2.3.13, it is essential to identify a suitable reference gene. A literature review revealed very little in the way of suitable blood-based reference genes so as well as the putative reference genes commonly used in other tissue types (β-actin, GAPDH, RPL19), a new reference gene was essential for determining an accurate $2^{-\Delta\Delta C_T}$ comparison of transcript expression levels.

Since there was a wealth of high throughput sequencing data available, it was pertinent to check which transcripts were most stable between the two strains in the RNAseq data. To achieve this, a coefficient of variance (CV) was calculated for each gene across all of the normalised counts for each sample (Figure 4.2).

$$Coefficient\ of\ Variance(CV\%) = \frac{Standard\ Deviation}{Mean}$$

Figure 4.2: Equation for calculating the Coefficient of Variance across all reads of any given gene

Genes were then ranked according to the lowest CV to give a final list of genes that show the lowest amount of variation across all of the samples from both strains (Table. 4.1). Also displayed are some of the common reference genes employed by the laboratory for various tissue types.

Gene	Coefficient of Variance (CV)
Pink1	*7.05%*
Slc25a39	*7.55%*
Abhd4	*7.59%*
Rnf10	*7.66%*
Oaz1	***7.79%***
Gapdh	*22.27%*
β-Actin	*16.46%*
RPL19	*33.47%*

Table 4.1: RNAseq Generated coefficient of variance (CV) as calculated via the equation outlined in the equation in Figure 4.2 from the n=6 cohort RNAseq experiment. Calculated CV values were then ranked from smallest to largest. Here Oaz1 (Ornithine decarboxylase antizyme 1) is highlighted as the reference gene selected for further validation. Also featured are the corresponding CV values for common putative reference genes.

4.4 Results

4.4.1 Validation of Findings in a New Cohort of Animals

qPCR Validation of Novel Reference Genes

Despite normalisation of total input RNA concentration during the cDNA synthesis stage of sample preparation, a clear difference in expression levels exist across either the strains or age groups for both GAPDH and β-Actin (Figure 4.3). β-Actin saw the

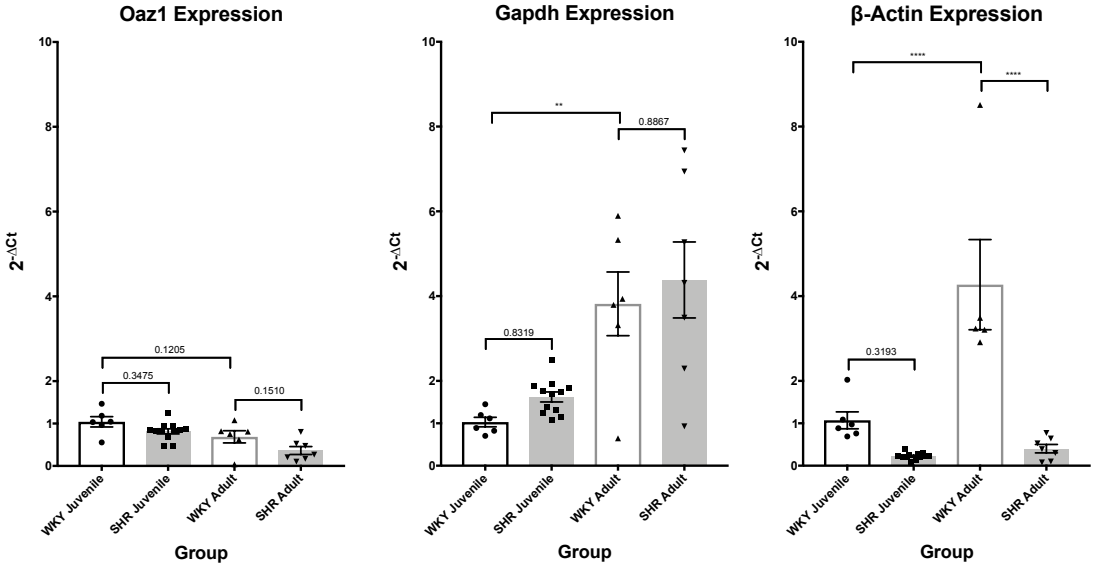


Figure 4.3: qPCR Blood Expression of Putative reference genes and the novel Oaz1 reference gene across Wistar Kyoto (WKY) and spontaneously hypertensive rat (SHR). Expression is displayed here as $2^{-\Delta C_T}$ normalised values for GAPDH, β-Actin, and the novel Oaz1 reference gene, in addition to the standard error of the mean (SEM) for juvenile and adult strains. Statistical analysis made use of two-way analysis of variance (ANOVA) with Tukey's multiple test correction. (P-Value <0.05, *; <0.01, **; <0.001, ***).

greatest level of change across the age groups, and across the comparison of Wistar Kyoto (WKY) *versus* spontaneously hypertensive rat (SHR) at the adult age group (Figure 4.3; juvenile WKY vs SHR, -0.85 fold, *P-Value=0.32*; adult WKY vs SHR, -3.87 fold, *P-Value<0.0001*; WKY juvenile vs adult, -3.2 fold, *P-Value<0.0001*).

The expression levels of GAPDH were much greater in the adult cohort as compared with the juveniles. However, there was no significant difference between the two strains at any given age. Furthermore, the juvenile cohort showed a much tighter spread of expression values, as compared with the spread observed within the adults (Figure 4.3; juvenile WKY vs SHR, +0.59 fold, *P-Value=0.83*; adult WKY vs SHR, +0.58 fold, *P-Value=0.89*; WKY juvenile vs adult, +2.79 fold, *P-Value=0.001*).

The levels of expresion observed for the Oaz1 appear much more stable across the experimental groups after a two-way ANOVA (Figure 4.3). Amending the scale for the Oaz1 expression plot shows the levels of variation between the groups (Figure 4.4). There appears a lower level of expression within hypertensive or aged animals, however this is

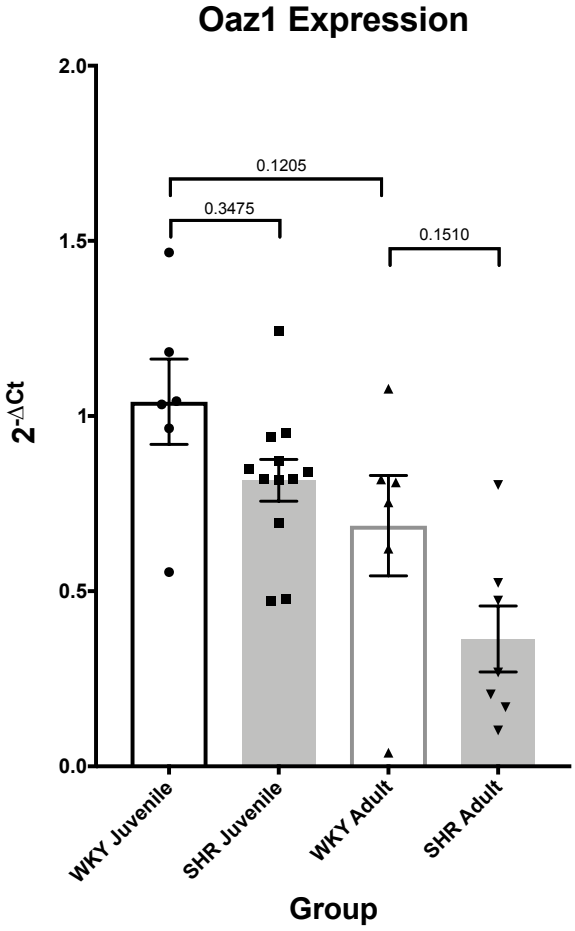


Figure 4.4: Rescaled qPCR Blood Expression of Oaz1 reference gene across Wistar Kyoto (WKY) and spontaneously hypertensive rat (SHR). Expression is displayed here as $2^{-\Delta C_T}$ normalised values for Oaz1 reference, in addition to the standard error of the mean (SEM) for juvenile and adult strains. Statistical analysis made use of two-way analysis of variance (ANOVA) with Tukey's multiple test correction. (P-Value <0.05, *; <0.01, **; <0.001, ***).

not significant in the juvenile or adult cohorts, as a high level of intra-group variation exists (Figure 4.4; juvenile WKY vs SHR, -0.22 fold, *P-Value=0.35*; adult WKY vs SHR, -0.32 fold, *P-Value=0.15*; WKY juvenile vs adult, -0.35 fold, *P-Value=0.12*).

qPCR Validation of Ifit1 in Brazilian cohort of Animals

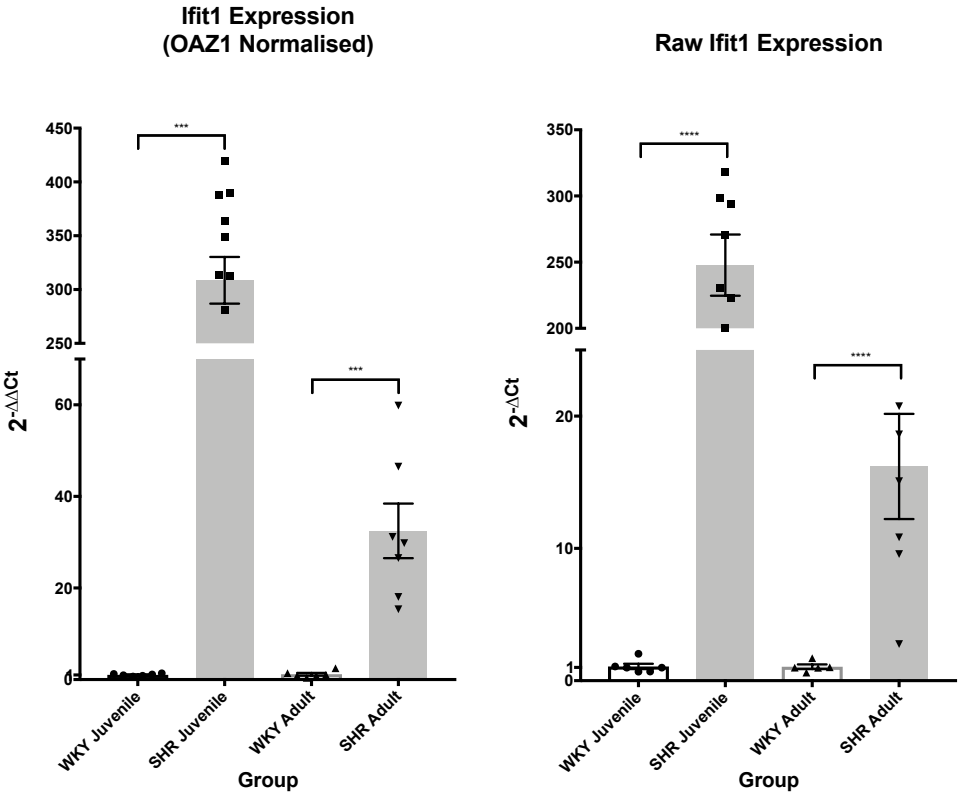


Figure 4.5: qPCR blood expression of Ifit1 across new cohort of Wistar Kyoto (WKY) and spontaneously hypertensive rat (SHR) animals sourced from Multidisciplinary Center for Biological Research (UNICAMP - São Paulo, Brazil). Here expression is displayed as both Oaz1 normalised $2^{-\Delta\Delta C_T}$ values and raw (unnormalised) Ifit1 $2^{-\Delta C_T}$ values. Statistics are presented as Two-way ANOVA P-values, between strains at any given age (P-value<0.05, *; P-value<0.01, **; P-value<0.001, ***).

qPCR analysis of the Ifit1 expression showed a similar, and hugely robust, profile of change as within other animal cohorts sourced from other suppliers (Figure 4.5). Here the strain differences were much more severe at both ages. This observation was further exagerated when normalised to the newly identified Oaz1 reference gene than when not normalised at all. Regardless of normalisation strategy, the differences in

Ifit1 blood expression are compelling within the Brazilian cohort of animals (Figure 4.5; juvenile WKY vs SHR, -0.22 fold, *P-Value=0.35*; adult WKY vs SHR, -0.32 fold, *P-Value=0.15*; WKY juvenile vs adult, -0.35 fold, *P-Value=0.12*).

4.4.2 Physiological metrics

Blood Pressure

Direct physiological measurements revealed a series of novel and expected patterns across the experimental groups (Figure 4.6). A clear and statistically significant difference in MAP exists between the adult WKY and SHR animals, as to be expected from this well established comparison (*P<0.0001*). The comparison between the juvenile WKY and SHR animals showed the SHR to have a significantly elevated blood pressure as compared with their normotensive and age-matched WKY controls (*P<0.0001*).

Spectral Analysis

Here spectral analysis revealed a significant increase in the LF domain of the BP spectra within the SHR animals compared with their normotensive controls in the adult cohort. This profile was consistent between SBP and DBP (Figure 4.7, SBP, juvenile WKY vs SHR, *P=0.25*; adult WKY vs SHR *P<0.0001*, DBP, juvenile WKY vs SHR, *P=0.08*; adult WKY vs SHR, *P<0.0001*). Furthermore, a significant increase is observed within the LF/HF ratio of the HR for the adult cohort only, pointed to an increased sympathetic output to the heart (Figure 4.7, juvenile WKY vs SHR, *P=0.85*; adult WKY vs SHR *P=0.002*).

Baroreceptor Sensitivity

Results from the analysis of the baroreceptor set point showed a clear increase in the SHR *versus* the WKY across both ages tested (Figure 4.8). Data points were clustered tightly and therefore resulted in a robustly significant P-Value from the Student's T-Test

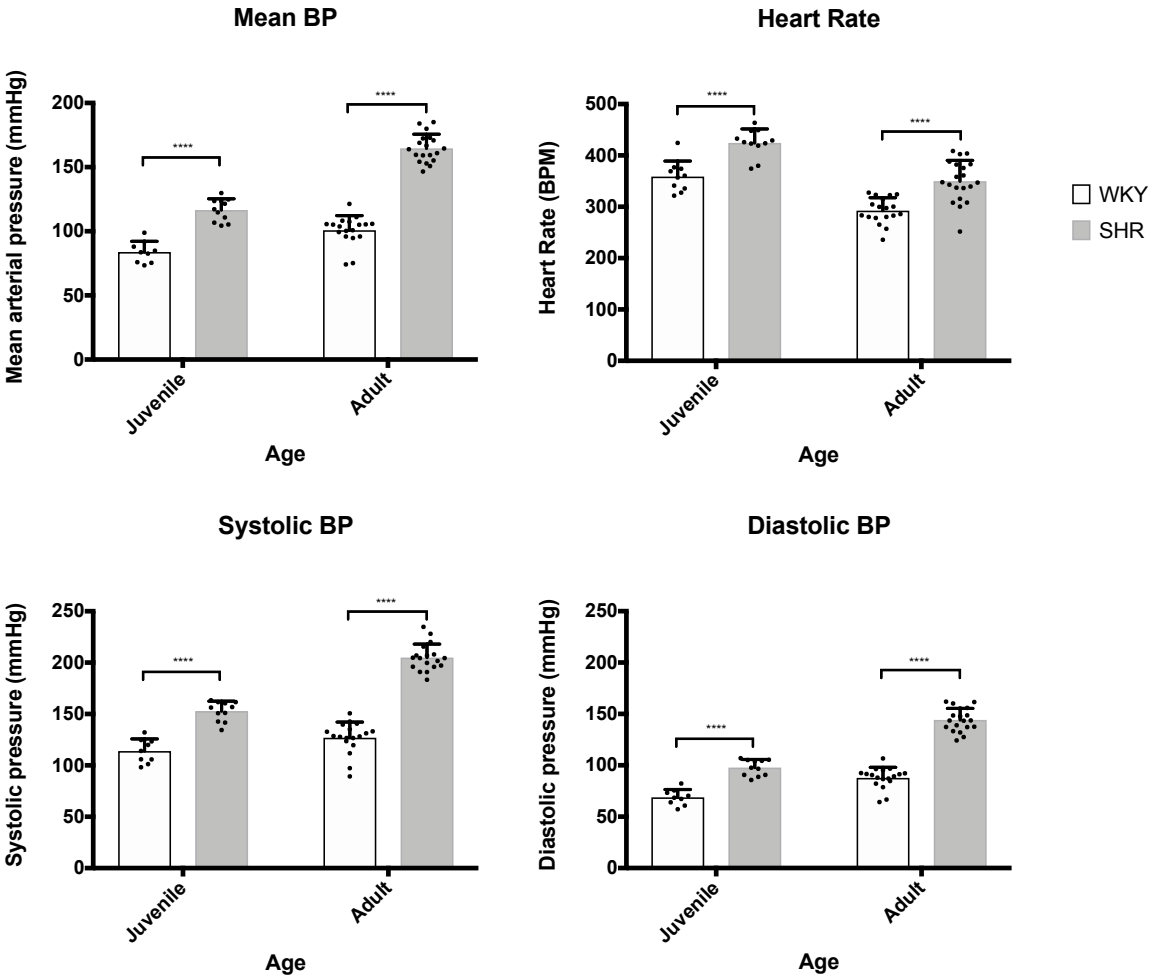


Figure 4.6: Physiological data in SHR and WKY strains at both juvenile and adult ages. Statistics are presented as unpaired Student's T-Test P-values, between strains at any given age (P-value<0.05, *; P-value<0.01, **; P-value<0.001, ***).

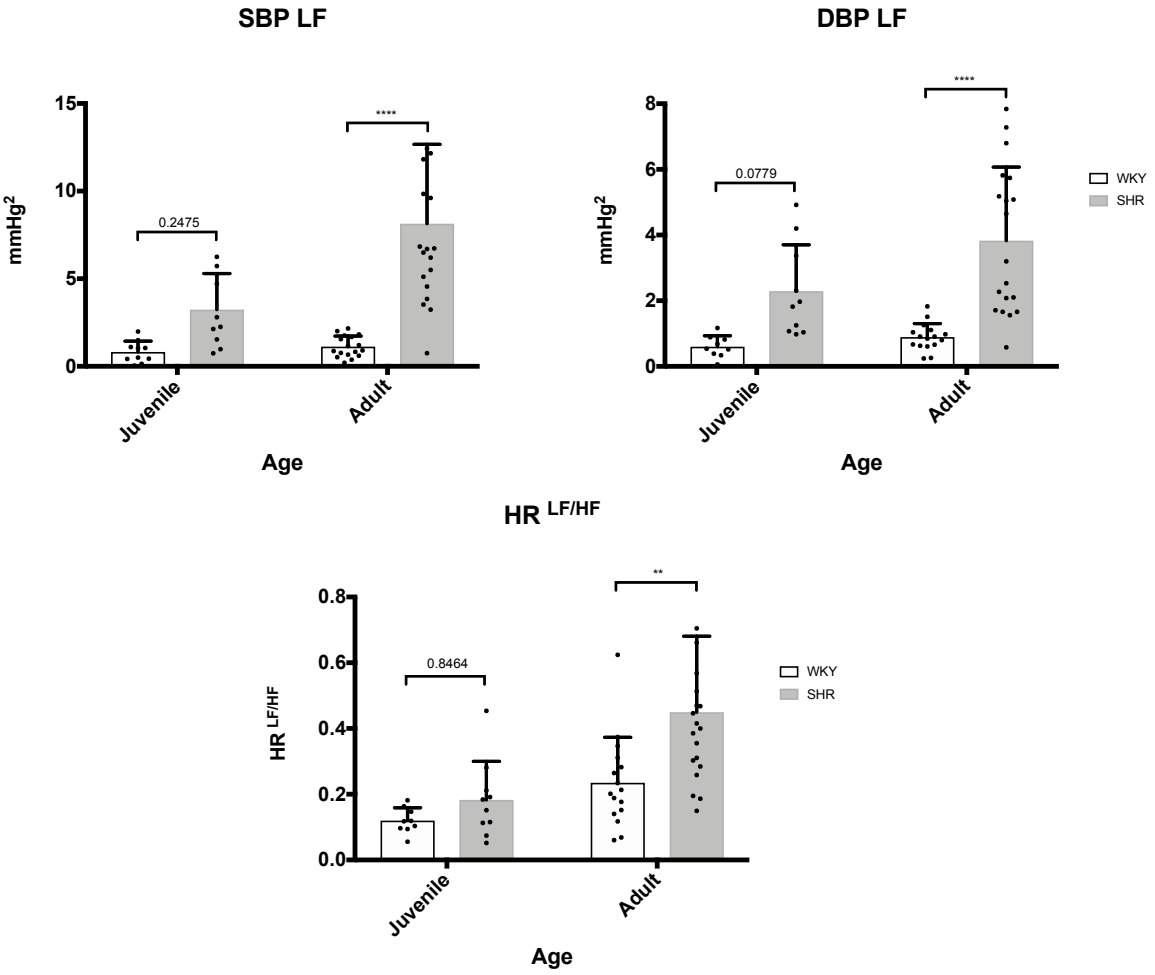


Figure 4.7: Spectral Analysis for WKY *versus* SHR comparison. Statistics are presented as unpaired Student's T-Test P-values, between strains at any given age (P-value<0.05, *; P-value<0.01, **; P-value<0.001, ***).

analysis (juvenile WKY vs SHR, *P<0.0001*; adult WKY vs SHR *P<0.0001*). Analysis of baroreceptor sensitivity saw a similarly robust difference, with SHR animals seeing a significantly reduced baroreceptor sensitivity compared to the WKY animals (Figure 4.8). This was consistent across both ages (juvenile WKY vs SHR, *P=0.0003*; adult WKY vs SHR *P<0.0001*).

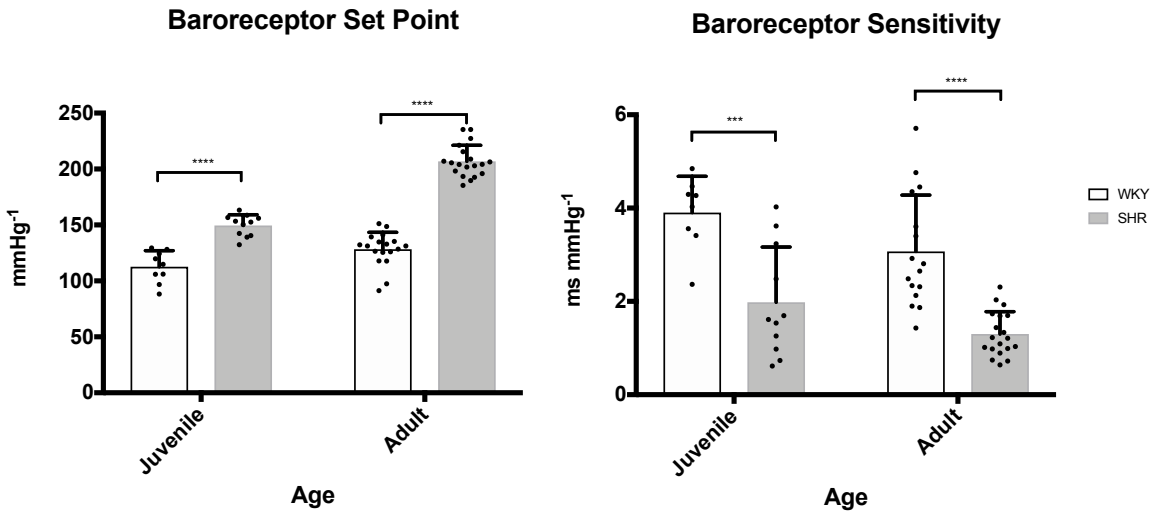


Figure 4.8: Baroreceptor Sensitivity ($mmHg^{-1}$) and Set Point ($mmHg^{-1}$) for WKY *versus* SHR comparison. Statistics are presented as unpaired Student's T-Test P-values, between strains at any given age (P-value<0.05, *; P-value<0.01, **; P-value<0.001, ***).

Correlation with Ifit1 expression levels

In order to deduce whether Ifit1 expression levels could be used as a direct measurement of blood pressure, linear regression modelling was employed using the Oaz1 normalised $2^{-\Delta\Delta C_T}$ values. Global regression modelling revealed little correlation between Ifit1 expression and MAP, with a low R^2 value and relatively neutral slope coefficient (Figure 4.9). Due to the distinct groupings being observed on the scatter graph, this analysis was repeated on subsets of the data, taken for either WKY, SHR, juvenile or adult.

The SHR subset saw by far the greatest R^2 value, denoting the best "Goodness of Fit" for the regression model (Figure 4.9; SHR, R^2 = 0.77, Slope = -0.16 (95% CI ± 0.02); WKY, R^2 = 0.03, Slope = -10.42 (95% CI ± 21.46)). However, as the juvenile

SHR animals had the highest expression of Ifit1 and lower MAP than the adult SHRs, this slope followed a negative coefficient of -0.16.

When the strains were separated based on age groups, a positive regression model was observed in both sets. However, both the R^2 and slope coefficients were relatively modest for these subsets (Figure 4.9; All adult, $R^2 = 0.57$, Slope = 1.32 (95% CI ± 0.38); All juvenile, $R^2 = 0.21$, Slope = 0.08 (95% CI ± 0.04)).

The linear regression analysis of Ifit1 blood expression against beats per minute (BPM) for all samples showed a stronger positive correlation. While the slope coefficient was modest, its accompanying R^2 value indicated a greater "Goodness of Fit" (Figure 4.10; $R^2 = 0.57$, Slope = 0.31 (95% CI ± 0.05)).

Splitting this regression between the WKY and SHR strains, the SHR again saw a much stronger R^2 value for its positive correlation (Figure 4.10; SHR, $R^2 = 0.57$, Slope = 0.26 (95% CI ± 0.06); WKY, $R^2 = 0.02$, Slope = -18.13 (95% CI ± 48.91))

Again, samples were further split into juvenile and adult subsets to assess whether Ifit1 saw a greater correlation within a specific age group. The resulting regression results were broadly comparable in terms of their R^2 values, however the adult cohort saw a subtle increase in the slope coefficient of the regression model (Figure 4.10; All adult, $R^2 = 0.53$, Slope = 1.67 (95% CI ± 0.52); All juvenile, $R^2 = 0.60$, Slope = 0.24 (95% CI ± 0.05)).

Baroreceptor sensitivity was plotted against blood expression of the transcript. Making use of all datapoints, from both ages and strains, regression modelling revealed little relationship between the two metrics (Figure 4.11; $R^2 = 0.04$, Slope = -0.005 (95% CI ± 0.004); SHR, $R^2 = 0.09$, Slope = 0.003 (95% CI ± 0.003); WKY, $R^2 = 0.0007$, Slope = 0.19 (95% CI ± 2.71); All adult, $R^2 = 0.36$, Slope = -0.04 (95% CI ± 0.02); All juvenile, $R^2 = 0.24$, Slope = -0.01 (95% CI ± 0.006)). All of the various subsets taken saw a low level of "Goodness of Fit", aside from the adult subsection which had the highest R^2 value of 0.36.

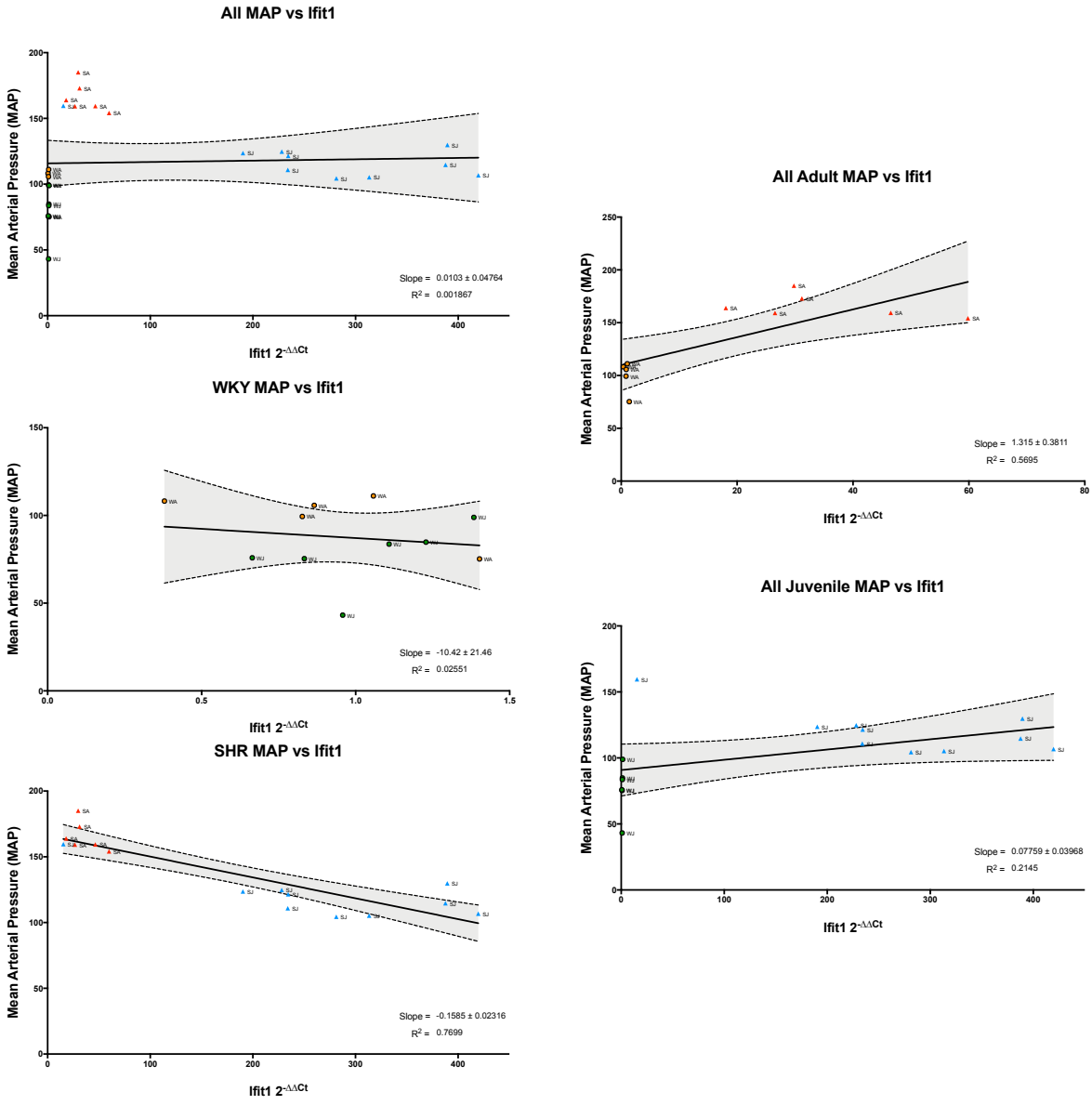


Figure 4.9: Linear Regression Analysis of Physiological mean arterial pressure (MAP), as millimetres of mercury (mmHg), against Oaz1 normalised Ifit1 blood expression $2^{-\Delta\Delta C_T}$ values. For the regression model, displayed here are the slope coefficient of the regression line and the R^2 correlation coefficient. Also highlighted in grey is the 95% confidence interval (CI) range for the model. Note the rescaled x-axis for the WKY plot.

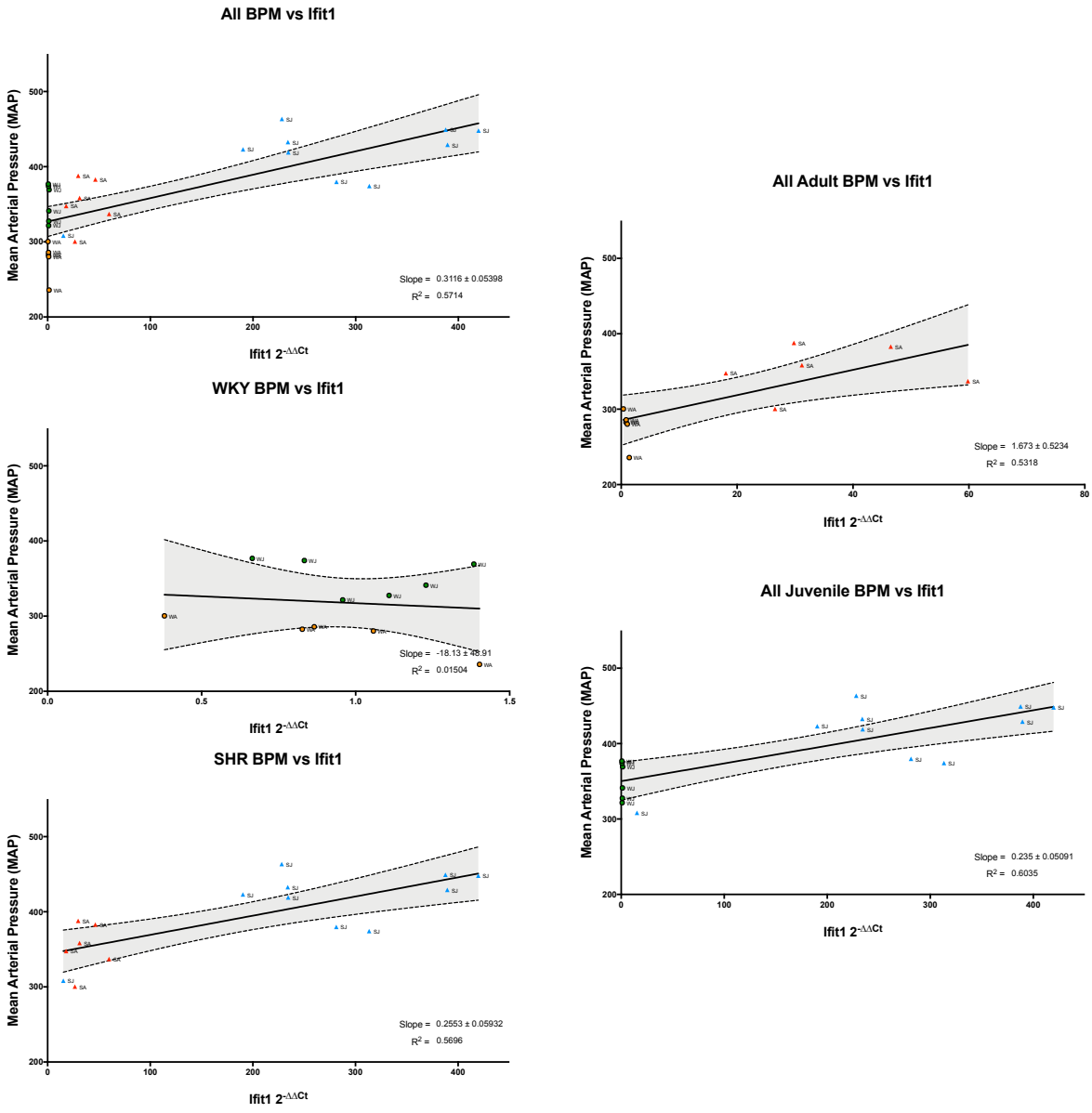


Figure 4.10: Linear Regression Analysis of Physiological beats per minute (BPM) against Oaz1 normalised Ifit1 blood expression $2^{-\Delta\Delta C_T}$ values. For the regression model, displayed here are the slope coefficient of the regression line and the R^2 correlation coefficient. Also highlighted in grey is the 95% confidence interval (CI) range for the model. Note the rescaled x-axis for the WKY plot.

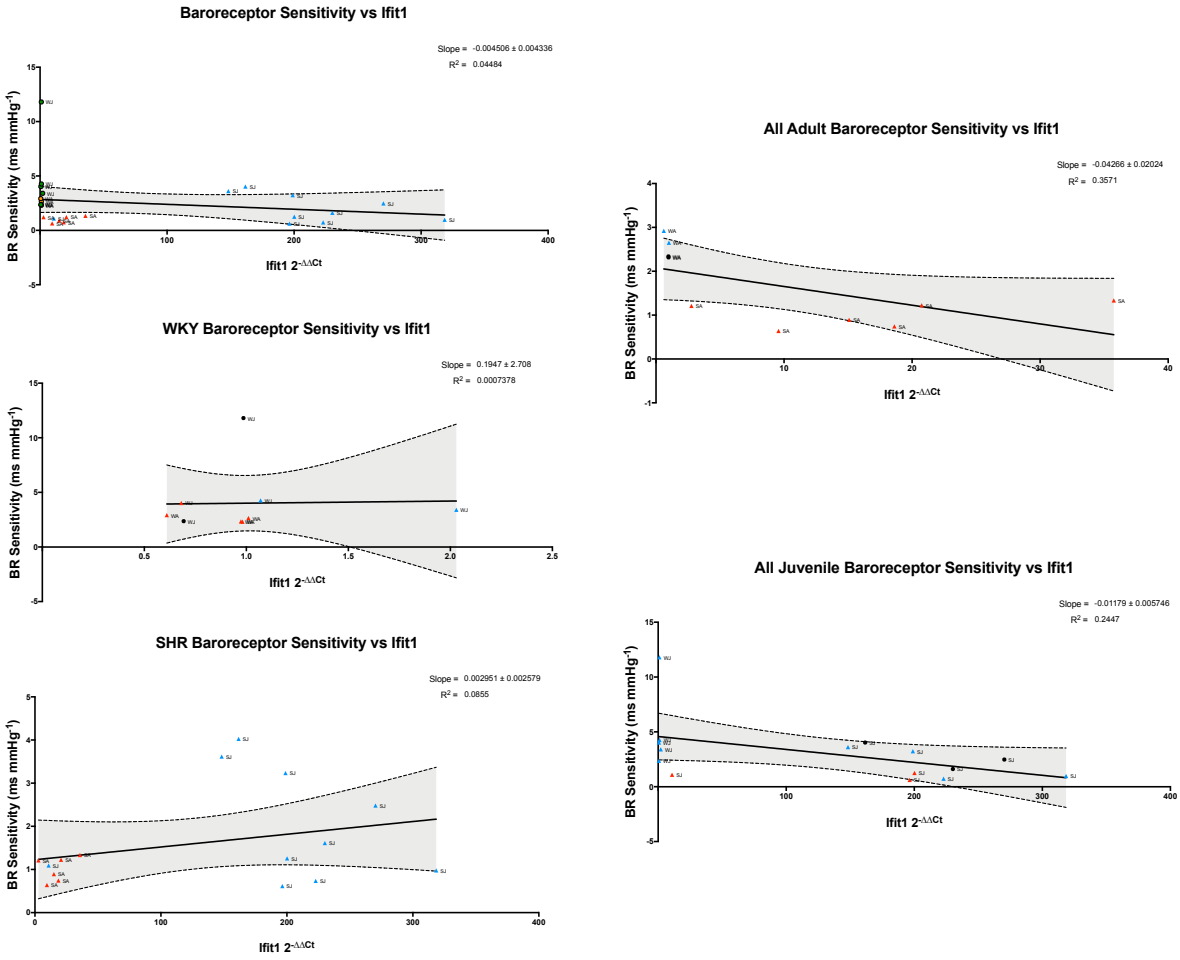


Figure 4.11: Linear Regression Analysis of Physiological baroreceptor sensitivity (BRS), as ms $mmHg^{-1}$, against Oaz1 normalised Ifit1 blood expression $2^{-\Delta\Delta C_T}$ values. For the regression model, displayed here are the slope coefficient of the regression line and the R^2 correlation coefficient. Also highlighted in grey is the 95% confidence interval (CI) range for the model. Note the rescaled x-axis for the WKY plot.

Linear regression modeling of Ifit1 against the Baroreceptor Set Point revealed little correlation when considering all animals of both ages, aside from the adult cohort (Figure 4.12; R^2 = 0.004, Slope = -0.02 (95% CI ± 0.07); SHR, R^2 = 0.69, Slope = -0.24 (95% CI ± 0.04); WKY, R^2 = 0.01, Slope = 4.04 (95% CI ± 13.14); All adult, R^2 = 0.64, Slope = 2.72 (95% CI ± 0.72); All juvenile, R^2 = 0.16, Slope = 0.09 (95% CI ± 0.06)).

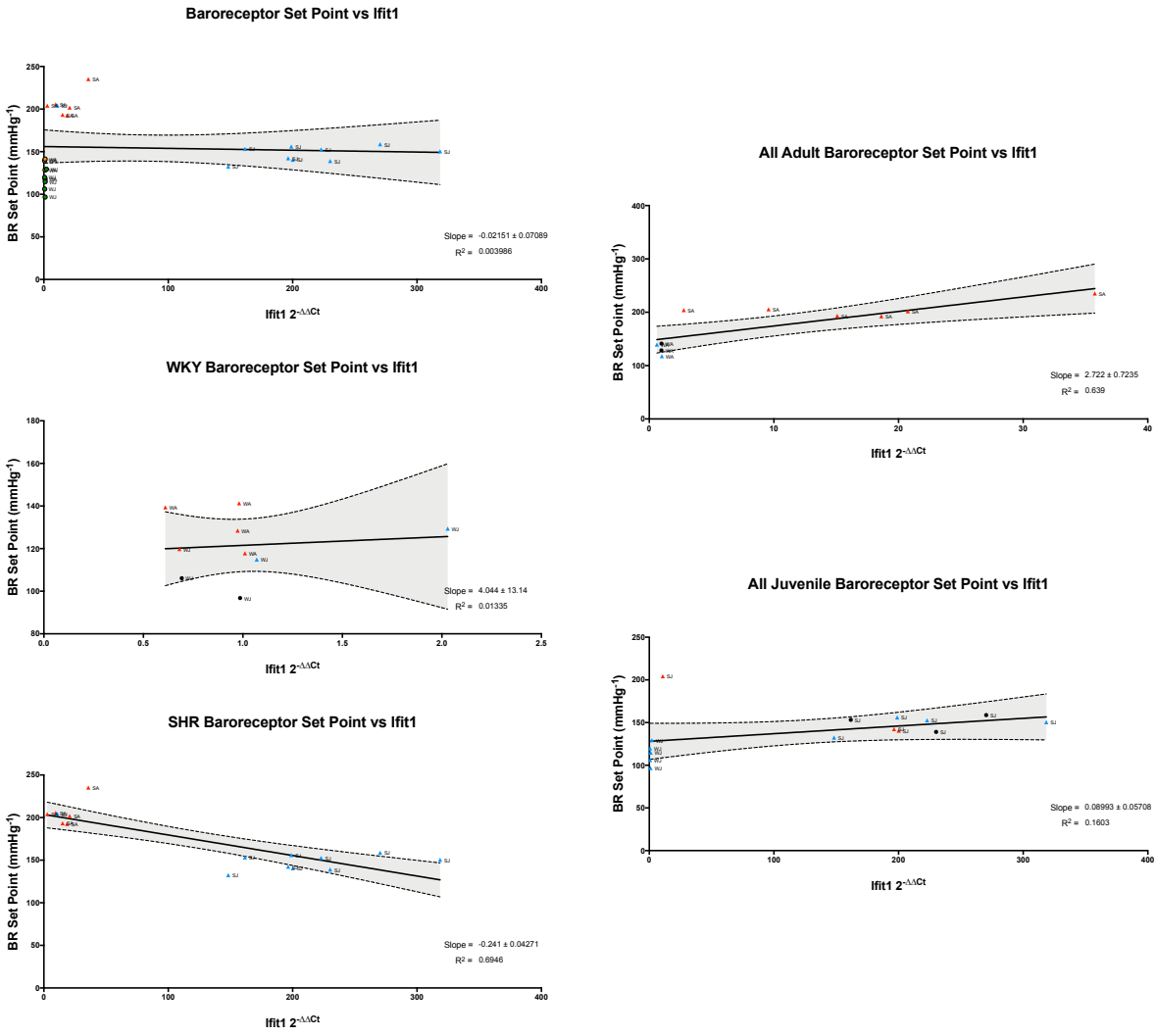


Figure 4.12: Linear Regression Analysis of Physiological Baroreceptor Set Point, as $mmHg^{-1}$, against Oaz1 normalised Ifit1 blood expression $2^{-\Delta\Delta C_T}$ values. For the regression model, displayed here are the slope coefficient of the regression line and the R^2 correlation coefficient. Also highlighted in grey is the 95% confidence interval (CI) range for the model. Note the rescaled x-axis for the WKY plot.

Metrics for spectral analysis of sympathetic activity saw very little correlation with Ifit1 expression in the blood. (Figure 4.13; SBP LF $R^2 = 0.004$, Slope = 0.003 (95% CI ± 0.008); DBP LF, $R^2 = 0.02$, Slope = 0.003 (95% CI ± 0.004); $HR^{LF/HF}$, $R^2 = 0.007$, Slope = -0.0001 (95% CI ± 0.0004)).

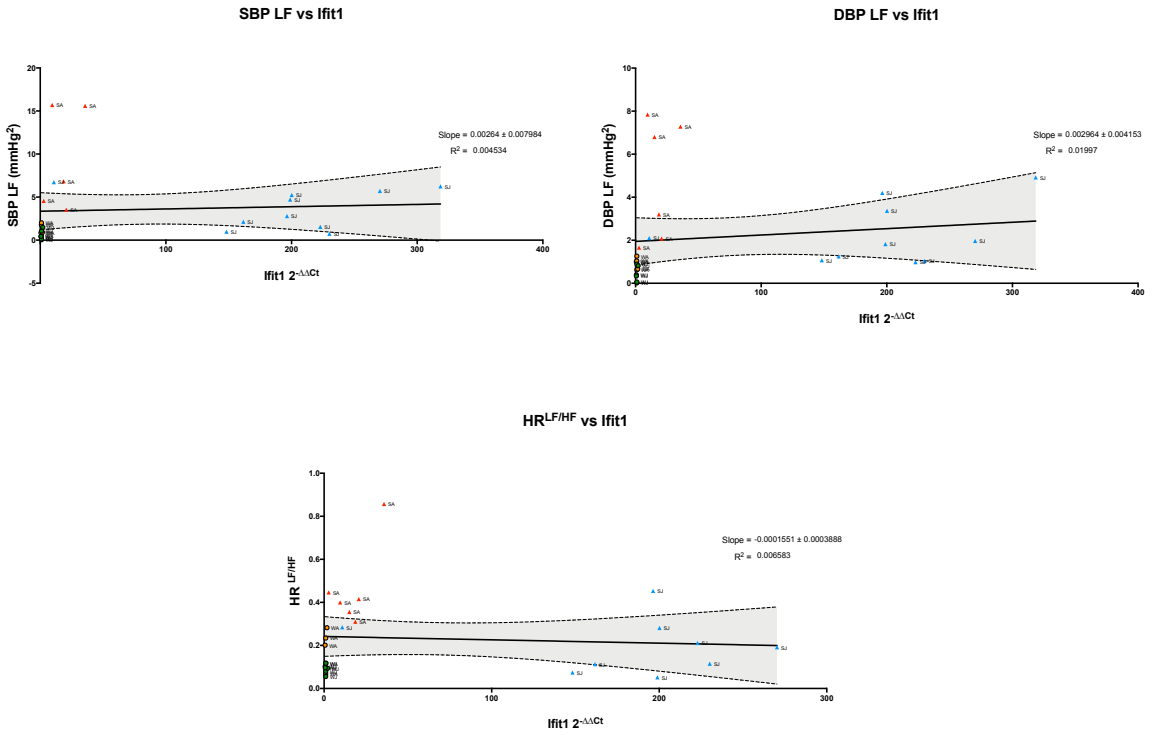


Figure 4.13: Linear Regression Analysis of Spectral Analysis Data, against Oaz1 normalised Ifit1 blood expression $2^{-\Delta\Delta C_T}$ values. Here Systolic blood pressure (SBP) and Diastolic blood pressure (DBP) Low Frequency spectra are used (LF; 0.27-0.74Hz, $mmHg^2$). Also displayed is the linear regression for heart rate (HR) ratio of LF to High Frequency (HF; 0.75 - 5Hz, $HR^{LF/HF}$). For the regression model, displayed here are the slope coefficient of the regression line and the R^2 correlation coefficient. Also highlighted in grey is the 95% confidence interval (CI) range for the model. Note the rescaled x-axis for the WKY plot.

4.5 Discussion

4.5.1 Methods of Obtaining Blood Pressure Readings

There exist three main methods of assessing blood pressure in a rat; radio telemetry, intra-arterial catheters, and tail cuff plethysmography. Each present their own advantages and disadvantages that must be considered for any experiment. Both radio telemetry and intra-arterial catheters are invasive methods of obtaining real time blood pressure readings, with intra-arterial yielding the most precise values.[147] The non-invasive tail cuff method requires animals to be handled and restrained for the period of measurement which may lead to falsely elevated blood pressure readings, as demonstrated in both mouse and rat studies.[148,149] Indeed, restraint stress has been used extensively as a method of inducing acute increases in blood pressure.[137] Therefore, and further to the additional precision, invasive blood pressure measurements are best suited for obtaining stabilised blood pressure readings, by mitigating the stress responses seen in the conscious rat.

4.5.2 Physiological Metrics of the SHR

Direct blood pressure measurements in juvenile animals revealed an interesting significant increase in MAP in the SHR compared with the WKY. This is contrary to the previous assumption that the 4-week old age presented a pre-hypertensive time point in the SHR. During the literature review conducted prior to the experiments outlined in section 2.4.2, multiple development profiles of when hypertension began were reported, ranging from 5-weeks to 7-weeks.[150,151] No papers could be found that directly measured blood pressure profiles across age specifically within the São Paulo cohort of SHR.

Analysis of the baroreceptor reflex showed a pattern consistant with the literature, for both baroreceptor set point and sensitivity. Multiple studies have noted the elevated set point within the SHR model of hypertension, and further observed a similar pattern

in humans exhibiting chronically elevated blood pressure.[137,152]The same is true with baroreceptor set point, with many using these key physiological metrics as a way of highlighting the suitability of the SHR model for human hypertension.[153,154] The observation that baroreceptor set point was higher in the SHR, and sensitivity was blunted amongst these hypertension animals, was therefore expected in this cohort of animals and well supported from the literature.

What is less clear from the literature, is the age at which these changes appear to occur in the SHR model. Here the juvenile cohort were 4 weeks old at the time of analysis. Considering the significant changes in blood pressure for these juvenile animals, it is not surprising that alterations are taking place to the baroreceptor response. However, changes in baroreceptor function have previously only been noted in animals as young as 8 weeks of age.[139,155,156]

This study made use of Spectral analysis derived LF spectra, in order to assess the sympathetic modulation of blood pressure regulation. For both the adult SHRs, SBP and DBP LF spectra were higher than in their age-matched controls. The observations that these metrics are significantly elevated in the SHR strain are well documented in the literature, as multiple studies have characterised the elevated sympathetic outflow in this model of hypertension.[39,132] Again, there does appear to be a disparity between the literature and these findings in terms of age. While sympathetic activation is known to precede the onset of hypertension, a rise in sympathetic outflow has only been noted in animals as young as 6 weeks.[157] Here physiological data shows the onset of hypertension as young as 4 weeks old and so an accompanying rise in sympathetic outflow is to be expected. However, a significant modulation towards sympathetic outflow was not observed in these animals through this analysis. For this reason it may be pertinent to attempt more direct sympathetic measurements rather than inferring from spectral readings.[158] With direct measurements we can validate spectral results as well as ascertain whether the juvenile SHR rats truly display elevated sympathetic activity or or not.

For these reasons, it may be necessary to take SHR animals at a younger age and compare them with age matched WKYs in order to truly study biomarker expression during the development of the disease. Currently, juvenile animals are obtained at 3-weeks of age and habituated on site for a week to ensure no confounding experimental effects due to elevated stress of shipment. This presents several potential problems however, notably the availability of animals younger than 3-weeks of age as suppliers do not recommend weaning prior to this age. Indeed, several studies have found direct effects of weaning on long-term physiological phenotypes.[159] These maternal deprivations can have prolonged effects on stress hormones critical in blood pressure regulation.[160,161] The disparities observed for age of onset of hypertension, and indeed the other physiological hallmarks that accompany it, may well be down to the genetic variation in the SHR strain, as explored later in section 4.5.4.

Correlation with Ifit1 expression levels

In an attempt to delve deeper into the possible suitability of Ifit1 expression as a metric of physiology, linear regression modelling was employed. Various metrics pertinent to blood pressure were assessed; mean arterial pressure, beats per minute, baroreceptor set point, and baroreceptor sensitivity. Regression modelling revealed a mixture of results.

One of the key reasons for the poor correlations between Ifit1 and the metric being assess, is the robust rise in Ifit1 levels in the juvenile animals. While the direct blood pressure measurements have shown the juvenile SHRs to have elevated blood pressure as compared with their age-matched controls, their blood pressure is not as fully developed as within the adult SHRs. This, coupled with the elevated expression of Ifit1 in SHR juveniles compared with SHR adults, leads to cluster of animals with modestly elevated blood pressure and maximum expression of ifit1. This cluster of juvenile SHRs throws the linear regression model and prevents an accurate regression line from being drawn.

With this in mind, an argument could be made for expanding upon this approach with the use of non-linear regression. Here for simplicity a linear model was employed.

However multiple biomarkers have been found to follow non-linear patterns in their diagnostic potential. For example, a study in 2011 found levels of cerebrospinal fluid β-Amyloid and Tau proteins to follow a biphasic relationship with development of the Alzheimer's disease.[162] For this reason, an iterative approach towards regression modelling may reveal a better "Goodness of Fit" for the model produced. However before this approach is employed, efforts need to be made to ensure the Ifit1 transcript is truly a suitable biomarker for hypertension and not simply due to a heterogeneous SHR transcriptome.

4.5.3 Oaz1 as a reference gene

A literature review of the genes showing the lowest CV percentages revealed that only Ornithine decarboxylase antizyme 1 (Oaz1) has been previously highlighted as a reference gene from a range of sample preparations.[163,164]

For this reason, Oaz1 was selected for validation against the new cohort of cDNA samples to confirm it was stable in its expression across both strains. Oaz1 performed suitably within the juvenile cohorts, as no significant difference existed between the strains given the equal quantity of RNA used for the analysis. This stability across the individuals therefore allows for each sample to be normalised against its expression.

As Oaz1 was selected based only RNAseq data from juvenile rats, there was no way of anticipating its expression profiles between the normotensive and hypertensive animals at the 12 week adult age. From the qPCR (Figure 4.4), Oaz1 showed a downward trend in the aged animals. This was not significant across ages within the same strain, however was between the WKY and SHR at the adult age. For this reason, oaz1 does not present a suitable reference gene for the adult comparison, preventing an accurate profiling of ifit1 expression as the animals age. In order to find a reference gene that does not display significant differences between the strains, at any given age, additional transcripts highlighted by the CV analysis should be assessed at the adult age.

An alternative to this, would be to carry out a high-throughput qPCR assay plate, such as those outlined in section 3.5.4. Indeed many of these plates have been designed for identifying suitable reference genes, and can be custom built to include the top reference gene candidates produced by the CV analysis outlined here. However often these can only analyse a single sample at a time, and once this is scaled to include multiple age groups across both strains, the cost becomes prohibitively large, and comparable to conducting RNAseq analysis on these groups.

4.5.4 Genetic spread of the SHR and WKY rats

The SHR is an inbred rat strain that was originally established from outbred Wistar rats that displayed elevated BP.[150] The normotensive control strain, the WKY, was established from the same Wistar stock as the SHR and as a result is commonly used as a control strain for the SHR for studies conducted around the world. Due to the global usage of the WKY and SHR strains, the sources of animals vary between studies. This translates to significant levels of variability in the results obtained and can be attributed, in part, to the fact that the WKY breeding stock was distributed to breeders before being fully inbred.[165–167] Furthermore, the littermates used for producing the inbred WKY from the Wistar rats were selected >10 years after the initiation of inbreeding for the SHR.[166] This further increases the heterogeneity between the strains, and has caused many to question the validity of using the WKY as a control for the SHR.[168]

While some substrains of the SHR have been observed, the level of homogeneity among SHRs is believed to be much greater than within WKYs. This is exemplified by the near 100% incidence of hypertension within the SHR across the world.[150,166] This observation is further supported here via direct blood pressure measurements in a colony of SHRs that were geographically remote from the animals used in section 2.4.2. A study in 2013 set out to map these differences in the genetic architecture of the WKY and SHR substrains from different sources.[127] The team took 16 commonly used WKY and

SHR substrains and used genome-wide SNP genotyping data to map levels of genetic drift. They found a large genetic divergence within both of the strains and were able to resolve the substrains through multidimensional scaling analysis (MDS), revealing distinct clustering based on the supplier from which the substrain was obtained. Based on the 9,407 SNPs covered, the team found a non-negligible genetic difference of 2.5% between two of the WKY substrains tested. They propose these subtle differences to be responsible for the significantly different phenotypes observed across behavioural paradigms from WKYs obtained from different vendors.[167]

With respect to the Ifit1 blood expression profiles observed for both the Brazilian and U.K. cohorts, a similar pattern of expression could be found. However, data from these experiments can only shed light on relative fold changes between the strains and can not be used to compare the cohorts themselves. These were isolated experiments, making use of relative abundances produced by internal normalisation, before being displayed as a ratio of expression between the strains. An interesting comparison could be made between the different colonies if tissue was obtained at the same time, and processed identically. However, this was not possible due to logistical difficulties when conducting experiments across the world. One way to mitigate this drawback would be to conduct an absolute assessment of transcript abundance, making use of a calibration curve displaying known quantities of Ifit1 transcript. This way, it would be valid to compare experiments that were conducted at different times or locations by plotting the absolute abundances of Ifit1. Given the variability of genetic architecture outlined above, an absolute metric of expression would show just how stable Ifit1 expression is across several cohorts of SHRs that all robustly develop hypertension.

4.5.5 Concluding remarks

Here and despite the varied genetic architecture of these strains globally, a similar expression profile of Ifit1 was observed within this geographically distinct cohort of

animals. This further validates its suitability as a comparative biomarker between these two strains.

Furthermore, this chapter was successful in clarifying an important assumption made at the outset of this work. Previous work focused on 4-weeks being a prehypertensive age within the SHR model of hypertension. Here direct blood pressure readings contradict this assumption, showing a significant increase in blood pressure between the WKY and SHR strains. This clarification is key, as it no longer allows for candidate biomarkers to be seen as prognostic at this age. However it does strengthen the use of Ifit1 as a generalised biomarker of hypertension as its profile remains elevated across these two ages tested.

Regardless, Ifit1 needs to be taken forward from the classical WKY *versus* SHR comparison and interrogated against additional animal models of hypertension. These additional animal models all vary in the aspects of human hypertension they emulate, potentially allowing a correlation between Ifit1 expression and the specific aetiology of the disease each model represents. By profiling its expression in additional animal models of hypertension, the transcript's suitability as a diagnostic tool can be determined.

Chapter 5

Widening the Search - Additional models of hypertension

5.1 Introduction

The profile differences observed in Ifit1 between the WKY and SHR strains are robust and significant, however there is a possibility these relative differences flagged up by qPCR and RNAseq could be attributable to the sole comparison of the WKY and SHR transcriptomes, and therefore not useful for identifying a hypertensive strain from amongst a population of other normotensive rat strains. In order to ascertain whether these biomarker expression profiles were truly indicative of the SHR hypertensive strain, a blood analysis across 3 additional rat strains was conducted. The 5 total strains selected for the expression profiling were therefore; spontaneously hypertensive rat (SHR), Wistar Kyoto (WKY), Wistar (WIS), Sprague-Dawley (SD) and Lister hooded (LH).

The key to this selection process was to provide a spectrum of normotensive strains that may act as controls to allow for hypertension specific transcripts to be identified. For this reason, a range of genetically similar and distinct strains were selected. In addition to the SHR and WKY strains previously focused on, the more genetically

distinct Wistar colony was also used from which both strains were originally obtained.[150]

The Wistar rat is an outbred albino rat that represents the first breed developed for use in biological and medical research. It was originally developed at the Wistar Institute. Prior to this, laboratories used the common house mouse (*Mus musculus*) as their primary model organism. More than 50% of all laboratory rat strains that are used today are descended from the original colony, developed in 1906.[169] The Wistar strain is one of the most popular rats used for research today, and has been further used to develop many common rat stocks such as the SHR and SD.

One of the more interesting strains added to this experiment is that of the SD. There exists a wealth of literature describing ways in which to evoke hypertension in this strain, be it surgically, pharmacologically or dietary.[170–172] It is therefore an appealing model to use for manipulating blood pressure levels using established methods and to study how blood pressure correlates with biomarker expression. Not only does this shed light on the correlation between blood pressure and biomarker expression, it also allows the reconciliation between the aetiology of hypertension and how transcripts are influenced. One of these methods is a chronic 3-day dehydration protocol. Previous studies have shown chronic dehydration to produce a significant increase in mean arterial pressure, from the second day of dehydration.[173]

An extension of this genetic diversity was given by the even more genetically distinct outbred LH rat strain. This was a strain originally bred from the Wistar outbred rat at the Lister Institute. Although commonly used in many experiments today the LH presented a strain that has seen little involvement in hypertension research.

In order to reconcile the interface between environmentally induced hypertension and genetically predisposed hypertension the borderline hypertensive rat (BHR) was later added to this series of experiments. This is a rat strain bred in order to study how a mixed genetic background can influence blood pressure. Lawler *et al.*, bred female SHRs with male normotensive WKY rats to produce a F1 generation offspring that had a genetic predisposition toward elevated blood pressure. The strain only exhibits a

slightly elevated blood pressure than WKY controls. However, when stressed or exposed to a diet high in salt, the strain develop overt hypertension at levels comparable to SHRs.[174,175] For this reason, this model presents an opportunity to study the complex relationship between the genome and environment in the aetiology of hypertension. Therefore, here an investigation on the expression profiles of biomarkers in the blood for stressed and unstressed animals at both pre- and fully hypertensive ages was conducted. The stress paradigm used was one established by Šarenac *et al.*, 2011 and consisted of a 9-day tube restraint course. This study replicated the stress paradigm in 4 and 12wk old animals.[137]

Another experimental method of inducing increased blood pressure is through the use of chronic dehydration. Several studies have observed this biological response to dehydration through the use of tail cuff or cannula based recording techniques of blood pressure.[176–178] A study in 2011 confirmed this by showing a 3-day dehydration protocol to robustly increase mean arterial blood pressure in a cohort of rats. The team go on to show the involvement of elevated sympathetic nerve activity acting upon its downstream effectors to raise blood pressure to a level higher than in euhydrated control animals.[16]

Using this many different control strains, from a range of different colonies, should allow for a better assessment of the suitability of the biomarker being tested. With such a large dataset of RNAseq significant transcripts between the WKY and SHR blood, any that fail to identify the SHR from the 4 control groups can be quickly and easily omitted without further study. Furthermore, characterising how genes change in response to an environmentally induced model of elevated blood pressure will help elucidate their potential role in hypertension.

5.2 Aims

The aim of this section is to take forward the work previously carried out and confirm the applicability of our previously identified biomarkers, in both genetically predisposed,

control and environmentally induced models of elevated blood pressure by;

- Assess Ifit1 expression using multiple control strains
- Use previously validated paradigms to environmentally induce acute blood pressure changes
- To use these previously characterised models of elevating blood pressure to act as proxy for studying transcript expression
- To profile any correlations between paradigms of inducing blood pressure changes and Ifit1 expression
- Establishing a BHR Restrained Vs Control Environmentally induced model of hypertension

5.3 Materials and Methods

5.3.1 Multi-strain Blood Analysis

Rats from each strain (SHR, WKY, Wistar, SD and LH; n = 8) were ordered at either 11 weeks of age or equivalent body weight based on supplier growth curves. This number of replicates per group was selected as a balance of sufficient power (as an increase from the n=3 and n=6 RNAseq experiments that showed significance for the Ifit1) and a reduction of the total number of animals required from the original qPCR experiments that made use of an n=12 per group. They were then left to habituate for 1 week before sacrifice via cranium strike and guillotine at 12wks. Trunk blood was taken and stored along with 2ml of RNAlater at -80°C before processing and qPCR analysis consistent with the above protocols. Brains were extracted and frozen on dry ice before storage at -80°C.

5.3.2 Sprague Dawley Dehydration

For the dehydration group, Adult (11wks) SDs were allowed to habituate for 1 week before removing water from their cages for 3 days (72 hours; onset 15:00 day 1, end 15:00 day 3). All DH rats had access to food. These were then sacrificed, along with a control group of experimentally naïve SDs (n=12/group), at 12wks & 3 days of age. Trunk blood and whole brains were collected and stored at -80°C before tissue processing and qPCR analysis consistent with the above protocols.

5.3.3 BHR Restrained Vs Control (4wks & 12wks)

Here a 9-day repeat restraint stress paradigm was implemented using the BHRs that could be compared to their age-matched Wistar controls. In addition to the stressed groups at both 4 & 12 weeks old, age matched control groups were set aside and received no stress. Animals in groups of 7 were housed at 25°C and supplied with food *ad libitum*. Juvenile and adult restraint group animals began the protocol 9 days prior to their 4th or 12th week of age respectively. The restrained protocol was the same as outlined in Šarenac *et al* 2011.[163] No direct, and potentially confounding blood pressure readings were taken, so this protocol was followed to enable the animals to act as proxies of elevated blood pressure. Briefly, each day at 9:00am, animals were habituated in the protocol room for 30mins before being placed for 60minutes in a clean plexiglass restraint (⌀ 3.8cm for juveniles, 5.5cm for adults) in the supine position. After which period, the animals were weighed. At the same time, the control group of animals were weighed daily. On the ninth and final day, animals were sacrificed immediately after restraint. Trunk blood and whole brains were collected and stored at -80°C before tissue processing and qPCR analysis consistent with the above protocols.

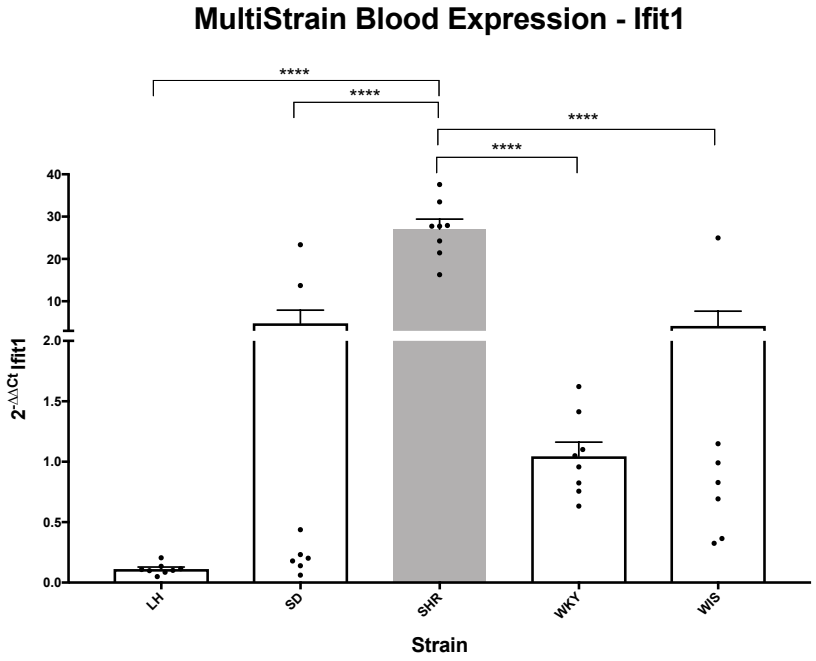


Figure 5.1: Multistrain Analysis of Ifit1 blood expression in 12-week old; Lister hooded (LH), Sprague-Dawley (SD), spontaneously hypertensive rat (SHR), Wistar Kyoto (WKY), and Wistar (Wis) rats. Displayed here are $2^{-\Delta\Delta C_T}$ of Ifit1 normalised to β-Actin and WKY expression. Statistical analysis made use of one-way analysis of variance (ANOVA) with Tukey's multiple test correction. (P-Value <0.05, *; <0.01, **; <0.001, ***; <0.0001, ****).

5.4 Results

5.4.1 Multi-strain Blood Analysis

Ifit1 expression analysis saw an interesting spread in blood expression across all of the various strains tested (Figure 5.1). Following a one-way ANOVA, and similar to the comparisons made in sections 2.4.2 and 4.4.1, Ifit1 saw a robust and significant increase in the SHR as compared with the WKY (Figure 5.1; WKY vs SHR, +26.02 Fold, *P-Value<0.0001*). This highly significant increase in abundance was conserved when comparing the SHR with any of the other control normotensive strains (Figure 5.1; LH vs SHR, +26.95 Fold, *P-Value<0.0001*; SD vs SHR, +22.27 Fold, *P-Value<0.0001*; Wistar vs SHR, +22.87 Fold, *P-Value<0.0001*;).

Despite being consistently elevated in the SHR, Ifit1 displayed a range of expression values within each of the multiple control strains tested. The LH blood saw by far the

lowest levels of Ifit1 expression, averaging at around 10% the expression values of the WKY. This was a tightly clustered grouping of expression, with little deviation from the mean (Mean=0.11, STDev=0.045). The SD cohort saw a much greater mean and spread due to 25% of the datapoints being greatly upregulated in expression to a level comparable with SHR expression (Mean=4.79; STDev=8.87). Here all expression values were normalised to WKY expression, as this had been the normotensive control for prior comparisons. Therefore, this became arbitrarily normalised as $2^{-\Delta\Delta C_T}$ values around 1.0. The spread of the WKY blood Ifit1 expression was relatively low compared with the other strains (Mean=1.04, STDev=0.33). Finally, Wistar Ifit1 expression showed a similar spread to the SD sample group, with a slightly diffuse profile of expression values, and a single outlier value on a scale comparable to those observed in the SHR blood expression values (Mean=4.19, STDev=9.17).

5.4.2 Sprague Dawley Dehydration

Following a 3-day chronic dehydration protocol, Ifit1 expression levels were assessed in the blood of SD rats and their euhydrated controls (Figure 5.2). Similarly to within the cohort of SD rats used for the multistrain analysis (Figure 5.1), Ifit1 saw a wide spread in expression values in the control euhydrated group. After 3-days of chronic dehydration, the spread of Ifit1 was much tighter than in the euhydrated cohort. Furthermore, an unpaired Student's T-Test revealed a highly significant downregulation in the dehydrated cohort as compared with the euhydrated controls (Figure 5.2; SD vs SD DH, -0.34 Fold, *P-Value<0.0001*).

5.4.3 BHR Restrained Vs Control (4wks & 12wks)

Expanding upon the dehydration method of induced elevated BP, a 9-day repeat restraint stress paradigm was initiated, based on Šarenac *et al.*[137] Blood expression of Ifit1 was assessed in both juvenile and adult rats (Figure 5.3). Statistical analyses took

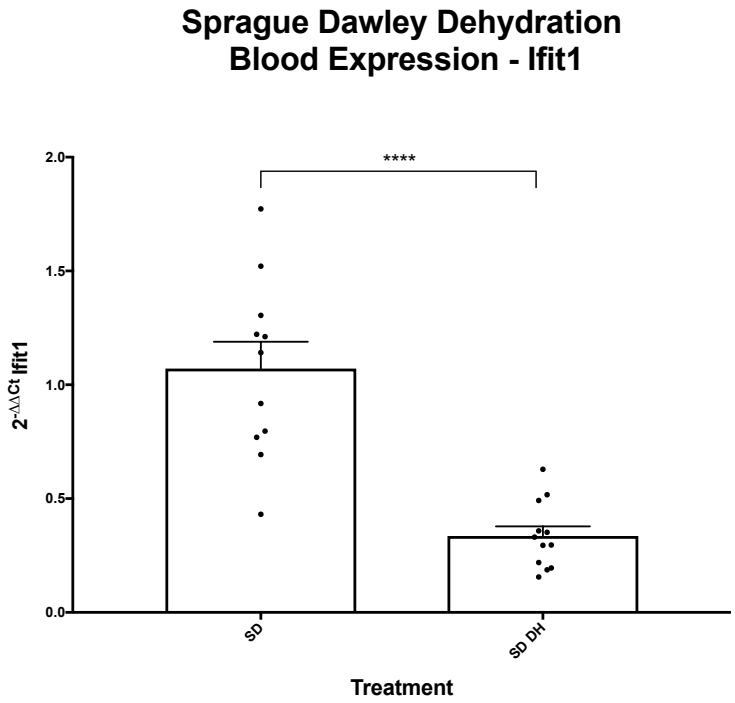


Figure 5.2: Assessment of Ifit1 expression levels in an environmentally induced model of hypertension; using 12 week old Sprague-Dawley (SD) rats and 3-days chronic dehydration (SD DH). Displayed here are $2^{-\Delta\Delta C_T}$ Ifit1 normalised to β-Actin and WKY expression. Statistical analysis made use of Student's T-Test. (P-Value<0.0001, ****).

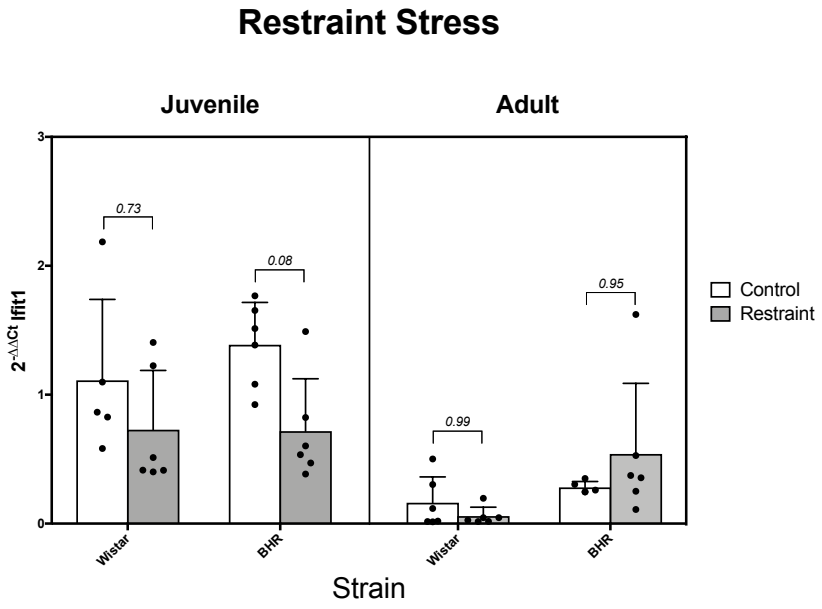


Figure 5.3: Restraint stress paradigm of induced hypertension, using genetically predisposed borderline hypertensive rat (BHR) against their Wistar controls. Both juvenile (4-weeks) and adult (12-weeks), were subjected to a 9-day repeat restraint stress paradigm or remained as unrestrained controls. Displayed here are $2^{-\Delta\Delta C_T}$ of Ifit1 normalised to β-Actin and control juvenile Wistar Ifit1 expression. Statistical analysis made use of three-way analysis of variance (ANOVA) with Tukey's multiple test correction. (P-Value <0.05, *; <0.01, **; <0.001, ***).

the form of a three-way ANOVA and showed no comparisons to reach below the 0.05 threshold of significance. Here the comparison of BHR control *versus* restrained in the juvenile cohort had the lowest P-Value of 0.08.

5.5 Discussion

5.5.1 Multi-Strain Analysis

The Multiple control strains were key to allowing the assessment of the suitability of candidate biomarkers. Using this method, a distinction could be drawn between a hypertensive strain, whilst ensuring any expression profile differences were not traits of the WKY control strain alone. As this body of work is looking for a biomarker that sheds light on individual susceptibility towards hypertension, it is to be expected that a successful biomarker would be tightly grouped amongst hypertensives (SHRs) and

spread amongst the normotensive controls. With this in mind, a genetically diverse set of control candidates were selected.

The genetic diversity required for this analysis has been confirmed and quantified through a variety of high-throughput studies. A study in 2013 looked into the genetic spread of 27 commonly used laboratory rat strains, including a variety of models for hypertension, insulin resistance, diabetes, and their resprective control strains.[179] Using next-generation illumina sequencing Atanur *et al* achieved between 10-20x read coverage, after rigorous filtering criteria, for all 27 strains. The team used the Brown Norway (BN/Mcwi) genome as a reference, and detected 9,665,340 single nucleotide variants (SNV) and 3,502,117 short indels. With this information, the team were able to calculate the distance between any pair of strains by dividing the number of differing SNVs by the length of the rat genome. A distance matrix could then be drawn using the Fitch-Margoliash method, with 1000 bootstraps. The resulting matrix was used to construct a phylogenetic tree (Figure 5.4). This independent analysis of rat strain heterogeneity illustrates the genetic diversity of the control strains selected here via the emergence of distinct clusters.

As this continuation was primarily a way of troubleshooting the ifit1 gene as a marker of hypertension, an additional and important consideration is the role ifit1 plays in the immune response. Ifit proteins, or Interferon-Induced proteins with tetratricopeptide repeats, are a major component of the protective host defences against viral infection encoded by interferon-stimulated genes (ISGs). Ifit genes are normally expressed at very low levels of silent all together and their expression can be triggered by many stimuli, although these are mainly within the context of viral and bacterial infection.[180] Therefore, it is important to work out whether the presence of elevated ifit1 coincides or precedes hypertension, or whether it is simply an artefact of a colony infection. For this reason, the decision was made to obtain animals from different colonies as well as different strains.

An interesting observation to note, was the separation of expression values that

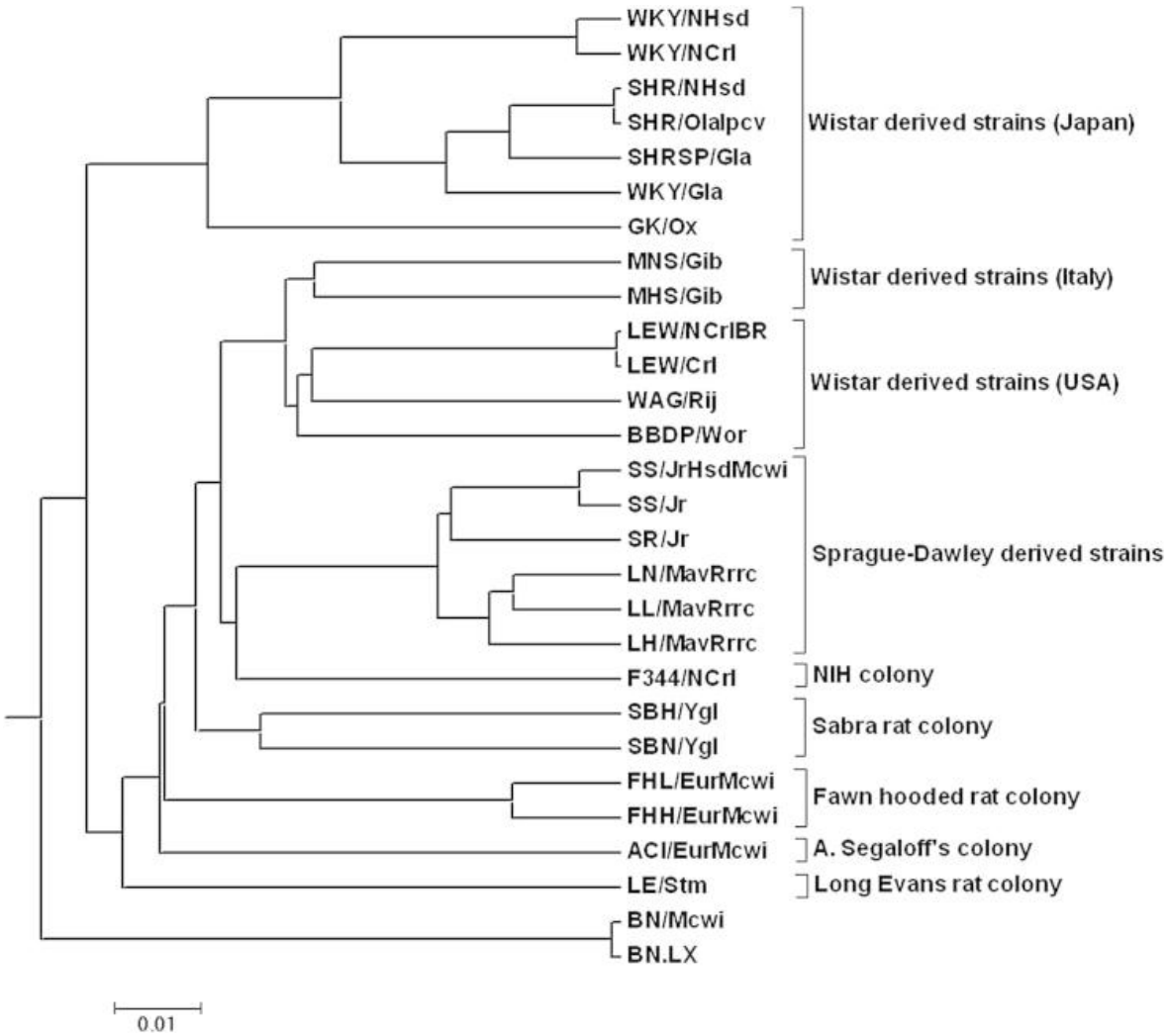


Figure 5.4: Phylogenetic Tree of 28 Rat Strains taken from Atanur *et al* 2013.[179] Over 9.6 million single nucleotide variants (SNV) were used to construct a phylogenetic tree with a scale representing genetic distance. This was calculated by dividing the number of single nucleotide variants (SNV) between a given pair of strains by the length of the reference genome (Brown Norway; BN/Mcwi). This tree was constructed using the Fitch-Margoliash method with 1,000 bootstraps. This tree illustrates the genetic diversity of the strains selected for further biomarker validation, amongst many more; Wistar Kyoto (WKY), spontaneously hypertensive rat (SHR), Goto-Kakizaki (GK), Milan normotensive strain (MNS), Milan hypertensive strain (MHS), Lewis (LEW), Wistar albino Glaxo (WAG), biobreeding diabetes prone (BBDP), salt sensitive (SS), salt resistant (SR), Lyon normotensive (LN), Lyon low blood pressure (LL), Lyon hypertensive (LH), Fischer 344 (F344), Sabra hypertension prone (SBH), Sabra normotensive (SBN), Fawn hooded low blood pressure (FHL), Fawn hooded high blood pressure (FHH), August & Copenhagen inbred (ACI), Long Evans (LE), Brown norway (BN).

appeared to stratify between a lower level of expression, and a greatly elevated expression, in an almost binary manner. This was exemplified in both the SD and Wistar groups, as both had outlier samples whose blood expression of Ifit1 was hugely elevated to an abundance comparable with the SHR levels of Ifit1 expression. This gave rise to several questions. Firstly, were these animals naturally occurring hypertensives/prehypertensives within their cohort. The strains with abundant Ifit1 expressing outliers were the Wistar and SDs. Both of which represent outbred strains with a greater level of genetic variability than the SHR or WKY strains. Indeed, the SHR was derived from Wistars exhibiting elevated BP.[150] It is conceivable that a true biomarker of hypertension, or a propensity towards hypertension, would show a natural variance in an outbred population, as this would be in line with the normal distribution of blood pressure. It was therefore hypothesised that individual biomarker expression profiles that outlie the normal range within their strain, and match the profile seen within the hypertensive strain, could be indicative of either pre- or fully hypertensive phenotypes. However, the only way to deduce this for certain would be with the addition of physiological BP data alongside blood transcript expression data, to see if a correlation could be drawn between these outliers and an elevated BP. Secondly, whether or not Ifit1 expression would provide a dynamic metric of blood pressure if elevated blood pressure were induced. Previous assessment of Ifit1 has been conducted in the genetically predisposed SHR strain, and not in an environmental model of induced hypertension. In order to answer both of these questions, and in lieu of direct blood pressure measurements, a well documented environmentally-induced model of hypertension was introduced on a new cohort of SDs in the form of a 3-day dehydration protocol.[173,181]

5.5.2 Sprague Dawley Dehydration

Based on the 5-strain analysis experiments, whereby 4 out of 12 (25%) SD animals displayed elevated ifit1 expression, one might expect 25% (n=4) of this new population to exhibit elevated ifit1. The naïve SD population saw a tight clustering of ifit1 expression values, with the majority similar to the SD population from the Multi-Strain Analysis (Figure 5.1). One of the rats showed a high expression of ifit1, comparable to both the 2 outliers of the Multi-Strain SD population, and the hypertensive SHR population. This was at a ratio lower than expected, as only one animal from a total of 12 exhibited this elevated expression profile. Interestingly, these expression profiles appear to again stratify between high and low expression, with none exhibiting expression levels in-between, as if ifit1 expression is switched on or off.

Following the 3-day dehydration protocol, Ifit1 saw a counter-intuitive reduction in abundance. In other words, an environmentally induced elevation of blood pressure occurred alongside a significant reduction in Ifit1 expression. The reasons behind this are difficult to deduce, as a literature review of ISGs and dehydration revealed very little work had been conducted on this subject. Indeed, high-throughput transcriptome analysis on male SDs from our own group has revealed a reduction in the total number of transcripts detected following a 3-day dehydration protocol.[182,183] However, these studies were all conducted in central regions of the brain tasked with cardiovascular and osmolarity control (supraoptic nucleus (SON); paraventricular nucleus (PVN); and neurintermediate lobe (NIL)) and not within whole blood. This reduction in abundance suggests Ifit1's inability to act as a dynamic metric of blood pressure when it is environmentally manipulated, however it does not belie its potential utility as a biomarker of genetically predisposition towards hypertension. One way of confirming this hypothesis further would be an analysis of its expression in the Dahl-Salt Sensitive model of hypertension.

The Dahl-Salt sensitive model of hypertension represents a strain with a genetic

predisposition towards hypertension. The strain can be split into two, based on their responsiveness towards a high salt diet; Salt sensitive and Salt resistant rats. Both of which steadily become hypertensive given a diet of standard rat chow. However, once given a diet high in salt, the salt sensitive cohort will develop a severe and fatal form of hypertension. As this model marries genetic predisposition (as seen in the SHR) with environmentally induced hypertension (as seen in the SD dehydration study), it may allow for the assessment of Ifit1 as a metric of the genome-environment aetiology of hypertension.

5.5.3 BHR Restraint

As the Dahl-Salt sensitive model was not available here, this work could be furthered in a similarly predisposed model by use of the borderline hypertensive rat (BHR). Due to its close relation to the SHR model, the BHR could be used to assess ifit1 expression in a similarly genetically predisposed strain, with only a modest elevated BP and in BHRs subjected to an environmentally induced form of elevated blood pressure.

The first thing noted by this study, was the high level of variability in the juvenile animals, regardless of strain or restraint. This high level of variation prevented any statistically significant calls being made by the three-way ANOVA employed here. Within the BHR group, the level of Ifit1 expression was more or less comparable to that observed within the Wistar controls. This is unlike the hugely robust fold changes observed within the SHR animals as compared with their normotensive controls. Indeed, there does appear a slight increase in expression in the BHR animals as compared with the Wistars, although this is not significant. A similar non-significant pattern is observed between the adult cohorts.

Assuming an elevation of blood pressure, the application of restraint stress appears to do little in modulating blood expression of Ifit1, regardless of the age group or strain tested. The juvenile BHR control *versus* restraint comparison has the lowest P-Value

(P-Value = 0.08), showing an almost significant reduction of expression.

Here the expression of Ifit1 was not significantly modulated in either control or restraint-stress treated animals. This was despite a model of induced elevated blood pressure within a genetically predisposed strain, that has been well documented to show an elevated BP.[137] Taken together with the results of the SD dehydration study, this is perhaps indicative of Ifit1's inability to serve as a biomarker of borderline or environmentally induced hypertension. In order to confirm this conclusion, the Dahl salt sensitive model would prove invaluable.

However, due to the genetic similarity between the BHR and SHR animals, it is promising to note a relatively low expression value in the control BHR. As this suggests the model destined to get hypertension (SHR) alone sees this over abundance of the transcript, and not the model of borderline hypertension.

5.5.4 Concluding remarks

This chapter was broadly successful in further characterising the ability of Ifit1 to discern between a hypertensive strain, and a cohort of normotensive controls. However, unlike the WKY *versus* SHR comparison which could be easily resolved by the stratification of highly or lowly expressed Ifit1, two of the control strains also saw elevated Ifit1 expression in the blood. While it was not possible to correlate elevated Ifit1 expression with an induced form of elevated blood pressure, Ifit1 may still serve as a suitable biomarker for assessing the distribution of blood pressure across outbred strains.

Since much of this work has been conducted via proxy animals following validated methods of inducing hypertension, future work should make use of direct physiological measurements in order to confirm a rise in blood pressure has truly taken place. By using plotting blood pressure changes, along with monitoring expression levels of Ifit1, a correlation can be drawn between the two. This way, Ifit1 suitability can be measured as a proxy metric of blood pressure variation, regardless of strain.

Chapter 6

Characterisation of Ifit1

6.1 Introduction

The natural progression after biomarker validation was to identify the role of interferon-induced protein with tetratricopeptide repeats 1 (Ifit1) within the body, and to ascertain whether Ifit1 had an involvement in any current mechanistic models of hypertension. Whilst a biomarker only needs to serve as a reliable and reproducable metric of blood pressure, and therefore does not need to be directly involved in hypertension, characterising the role of ifit1 in this disease may well reveal novel insights into its onset.

The Ifit family of proteins includes four canonical members within humans (Ifit1, Ifit2, Ifit3, and Ifit5), and is widely conserved in mammals, amphibians, and fish.[184,185] Within rats specifically there are 4 ifit proteins; Ifit1, Ifit1b, Ifit2, and Ifit3.[186] Interestingly, Ifit proteins cannot be found in lower order animals, such as the fruit fly (*Drosophila melanogaster*), nematode (*Caenorhabditis elegans*), yeast (*Saccharomyces cerevisiae*) or plants[187,188] (Figures 6.2 and 6.1).

Generally Ifit family genes are less abundantly expressed whilst in the absence of stimuli. However their expression rapidly increases during stimulation with IFN, viral infection, or other pathogen associated molecular pattern (PAMP) recognition. Most

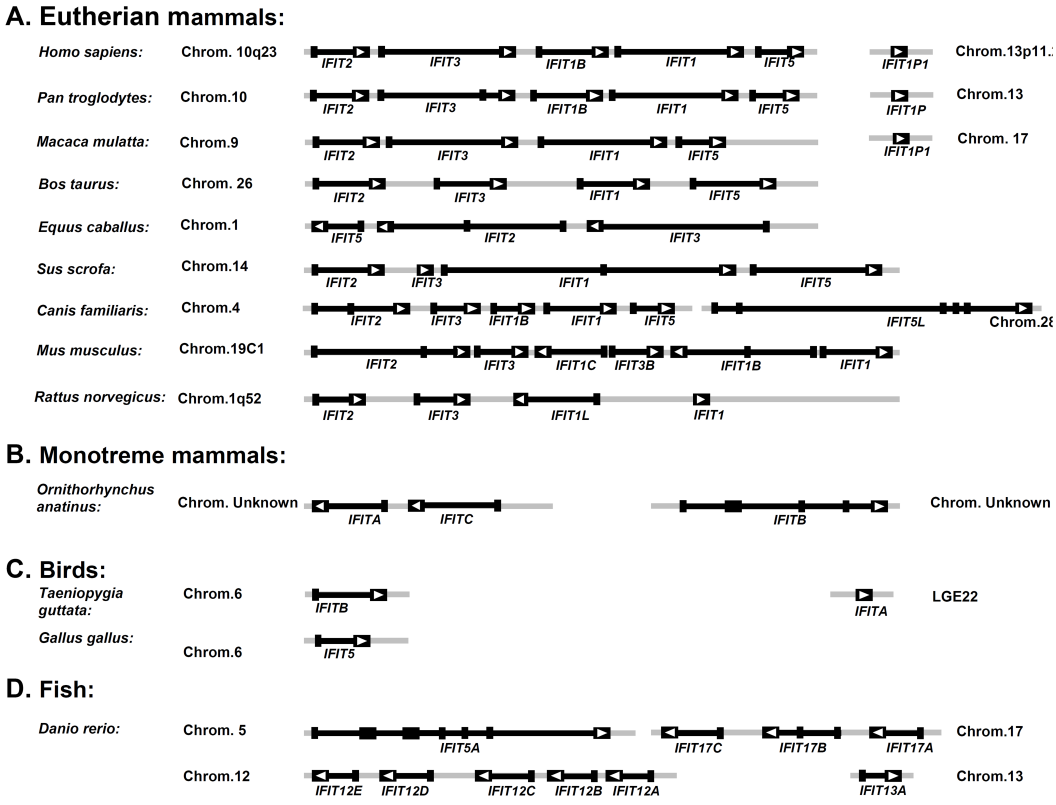


Figure 6.1: Exon-intron structural differences in Ifit family proteins, as taken from Liu *et al* 2013.[189] The exons are indicated by black boxes, and the introns by grey lines. The orientation of the open reading frames is indicated by arrowheads.

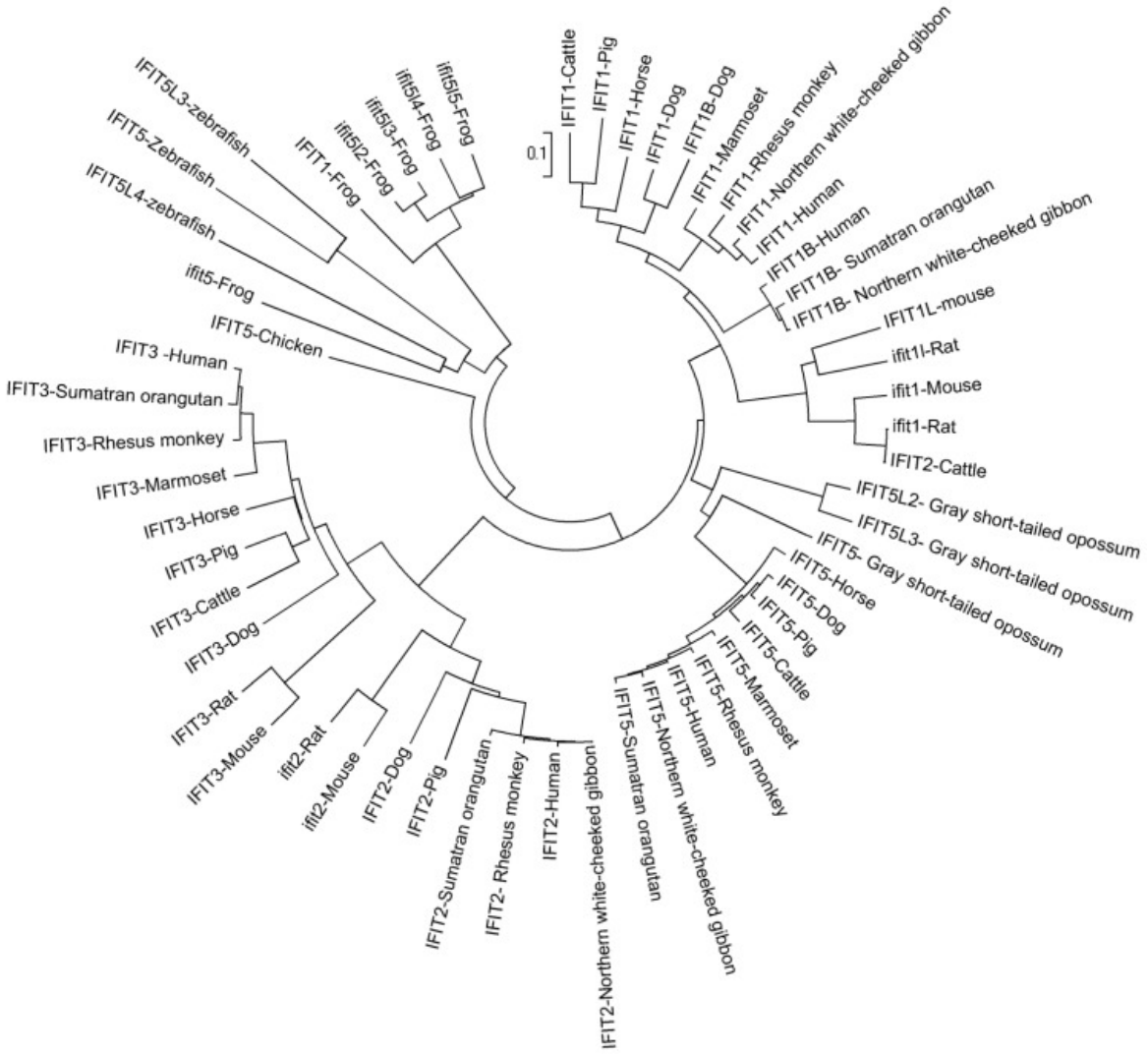


Figure 6.2: Phylogeny of Ifit family proteins, as taken from Zhou *et al* 2013, and generated by MEGA4.0[186, 190]

of the Ifit family of genes have two exons and contain between two or three interferon stimulated response elements (ISRE) as part of their promoter regions.[191] The presence of ISREs within the promoter of Ifits serves to explain their low basal expression levels and rapid IFN-dependent induction.[192] These ISRE are crucial cis-acting response elements that are recognised by IFN-Stimulated Gene Factor 3 (ISGF3), a factor that is induced by IFNs and various other stimuli.[193] The main inducers of the ifit genes therefore centre around the interferons (IFNs) and include; type I IFNs (IFN-α/β), type III IFNs (IFN-λ), and to a much weaker degree Type II IFNs (IFN-γ).[194,195]

The Ifit protein family are mostly described for their role in antiviral immunity. While none of the ifit family of proteins have any known enzymatic activity, they can inhibit replication of viruses by binding and modulating the functions of cellular and viral proteins and RNAs. All Ifits commonly share a tetratricopeptide (TPR) domain containing a 34 amino acid motif with the consensus sequence $[WLF] - X_{(2)} - [LIM] - [GAS] - X_{(2)} - [YLF] - X_{(8)} - [ASE] - X_{(3)} - [FYL] - X_{(2)} - [ASL] - X_{(4)} - [PKE]$. This motif folds into a resulting helix-turn-helix structure giving rise to a series of unique concave and convex surfaces that permit binding to a diverse range of ligands. TPR domains themselves are conserved in all kingdoms of life, and are considered to function as protein and peptide recognition sites. The discovery of Ifit's ability to bind RNA has served to broaden the known ligand spectrum of TPR motifs to include nucleic acids.

Through these TPR domains, ifit family proteins can modulate protein interactions.[196] For example, both Ifit1 and Ifit2 are involved in a nonspecific antiviral effect via their direct interaction with eukaryotic initiation factor 3 (eIF3) translation complex. The eIF3 complex is essential for translation initiation for several reasons; recruitment of mRNA, scanning of the mRNA for a start codon, and delivery of tRNA to the translation machinery.[197] Through binding with the subunits of the eIF3 protein, Ifits can block mRNA recruitment and tRNA delivery.[198] This leads to around 60% suppression of translation within the cell during protein synthesis which can indeed be detrimental

to the host cell, however this also prevents host-dependent viral replication.[199–201]

As well as inhibiting endogenous aspects of the viral replication pathway, recent evidence has shown the ability of Ifit family proteins to directly bind viral RNA. Generally speaking, host cytoplasmic RNAs are single stranded and mRNAs and rRNAs/tRNAs contain either a N-7-methylated guanosine or 5'-monophosphate cap, bound by a 5'-to-5' triphosphate bridge to the first base respectively. Certain higher order eukaryotes see a further modification to the mRNA in the form of a methylation of the first ribose.[202,203] These modifications confer more than simply translational regulation as their absence aids in the detection of foreign nucleic acids. Viruses in contrast may form double-stranded RNA, with the potential addition of 5'-Triphosphate modification during their life cycle. Viruses carrying this 5'-Triphosphate-RNA are recognised and bound by Ifit1, before an Ifit complex (containing Ifit1, Ifit2, and Ifit3) sequester the bound virus for destruction.[184]

While little has been published on Ifit1's potential role in hypertension, there is a great deal of potential for linking the two due to the well established link between hypertension and IFN expression. When focusing on the innate arm of immunity in hypertension, it has been established that angiotensin II infusions induce an increase in IFN-γ in the spleen and kidney of hypertensive rats.[204] Studies on IFN's role in inflammation may well provide the link between hypertension and ISGs such as the Ifit family of proteins.

6.1.1 Coupling Expression to Cardiovascular Centres of the Brain

As much of the work of this laboratory is centred around transcriptome analysis of central regions of cardiovascular control there have been terabytes of next-generation sequencing data generated, both published and unpublished. This presented a novel opportunity to reprocess these experimental outputs using the most up to date analysis

pipelines as optimised previously. The resulting datasets could be re-mined to assess whether Ifit1 expression was altered centrally between these strains, as it was within the blood. One of the most revealing experimental designs, was the comparison of WKY *versus* SHR at 12 weeks of age. The experiment isolated and sequenced three brain regions that are key in sympathetic and hormonal control of blood pressure, namely the; paraventricular nucleus (PVN), nucleus tractus solitarii (NTS), and rostral ventrolateral medulla (RVLM). Following a re-analysis by way of the optimised pipeline outline above, Ifit1 showed a significant upregulation in the SHR *versus* the WKY for both the PVN and RVLM (Table 6.1). While the fold change increase was relatively modest compared with the changes observed within blood, the read counts were consistent resulting in a relatively low P-Value. Due to the integrative nature of the PVN, as outlined in section 1.1.4, this region was selected for further validation to assess Ifit1 expression.

			DESeq2		
Region	Gene Name	WKY.AvgCount	SHR.AvgCount	FC	P-Value
PVN	Ifit1	23.73	67.88	1.63	2.77E-05
NTS	Ifit1	21.92	73.18	1.96	NA
RVLM	Ifit1	14.50	52.23	1.90	7.10E-05

Table 6.1: RNAseq generated counts for Ifit1 expression in cardiovascular centres of the brain. Based on n=3 experiments conducted on WKY *versus* SHR at 12 weeks of age, on cryostat isolated brain regions; paraventricular nucleus (PVN), nucleus tractus solitarii (NTS), and rostral ventrolateral medulla (RVLM). N.B. the "NA" P-Values issued by DESeq2 are returned due to a single sample with an extreme outlier (as detected by Cook's distance).

6.1.2 fluorescence-activated cell sorting (FACS)

Up until now this work has focused on Ifit1 expression in whole blood samples. However blood itself is comprised of several subpopulations of cell types, each with a different role in physiology. In order to delve further into characterising the overabundance of Ifit1 in the hypertensive (SHR) blood the specific subtype of blood cell responsible for its upregulation needed resolving. Of the major blood cell types (i.e. erythrocytes,

leukocytes, and thrombocytes), only leukocytes contain a nucleus (nucleated), and are therefore the focus of this study on transcription. Leukocytes as a type can be further split into several different subtypes, each with their own specific role in immunity.

With this in mind fluorescence-activated cell sorting (FACS) presented an ideal technology for resolving whole blood into its individual constituents, enabling Ifit1 expression to be quantified across each subpopulation of leukocytes. Cell sorting takes place based on the presence or absence of specific characteristics, in this case the presence of fluorophores conjugated to antibodies that can bind cell surface protein markers. A combination of excitation lasers and detectors enable the physical separation of cells, depending on how the operator calibrates the machine to isolate cells based on a combination of fluorophore emission (a process known as "Gating").

The difficulty presented with this technology, is the optimisation of antibodies to be used and the gating strategy to be employed. With this in mind a literature review revealed a team had optimised and validated a FACS pipeline for rat blood. In 2016 a 9-colour FACS method to characterise major leukocyte populations in the rat was carried out.[205] Barnett-Vanes *et al.* were able to produce a FACS pipeline that resulted in resolving the key blood leukocyte factions. The protocol made use of a comprehensive gating strategy to resolve the major blood leukocyte populations present in rats (Figure 6.3). In addition to this, the team went on to validate their approach using a model of lipopolysaccharide (LPS)-induced pulmonary inflammation. The team exposed Sprague-Dawley to LPS (1mg/mL), and collected several tissues including; blood, broncho-alveolar lavage fluid (BALF) and lungs. After having carried out a flow-cytometry separation, Enzyme-Linked Immunosorbent Assay (ELISA) was used on BALF tissue to validate their approach. Here the pipeline outlined in Barnett-Vanes *et al.* became a foundation for these experiments. It allowed for the resolution the various cellular populations contained within whole blood for the characterisation of which blood fraction was responsible for the robustly elevated Ifit1 expression in the SHR animals. By identifying this fraction it may be possible to guide a future literature review to

shed light on Ifit1 and its link towards an immune-based aetiology of hypertension.

6.2 Aims

In this chapter, and building upon the discovery of Ifit1 as a potential biomarker, a further characterisation of its expression and involvement in genetically predisposed hypertension will be conducted. This will be achieved via the following;

- Assess expression within a key cardiovascular centre in the brain
- Characterise the faction of blood responsible for the elevated expression of ifit1 in the SHR
- Conducting a literature review to assess role in mechanistic model of hypertension

6.3 Materials and Methods

6.3.1 Tissue Collection - PVN Isolation

Rats (WKY adults, SHR adults [12 Weeks], WKY juvenile, and SHR juvenile [4 Weeks] n=8 per group) were stunned by a blow to the head and then decapitated. A scalpel was used to make a longitudinal incision (rostral-caudal) on the crest of the head, and the scalp was pulled back to reveal the cranium. Two cuts were made, approaching from the site of decapitation to either side of the hindbrain using scissors. This enabled forceps to prise open the top of the cranium, allowing the brain to be gently liberated with a flat spatula. Brains were instantly frozen using powdered dry ice and stored at -80°C until needed. The brain was sectioned and mapped using a cryostat (Leica Microsystems CM1900 Cryostat). Coronal sections of tissue were cut rostral-caudal at 60μm thickness and stained with Toluidine blue (0.1% wt/vol in 70% EtOH; Sigma Aldrich) in order to map the hypothalamus (Paxinos and Watson Rat Atlas) (Figure

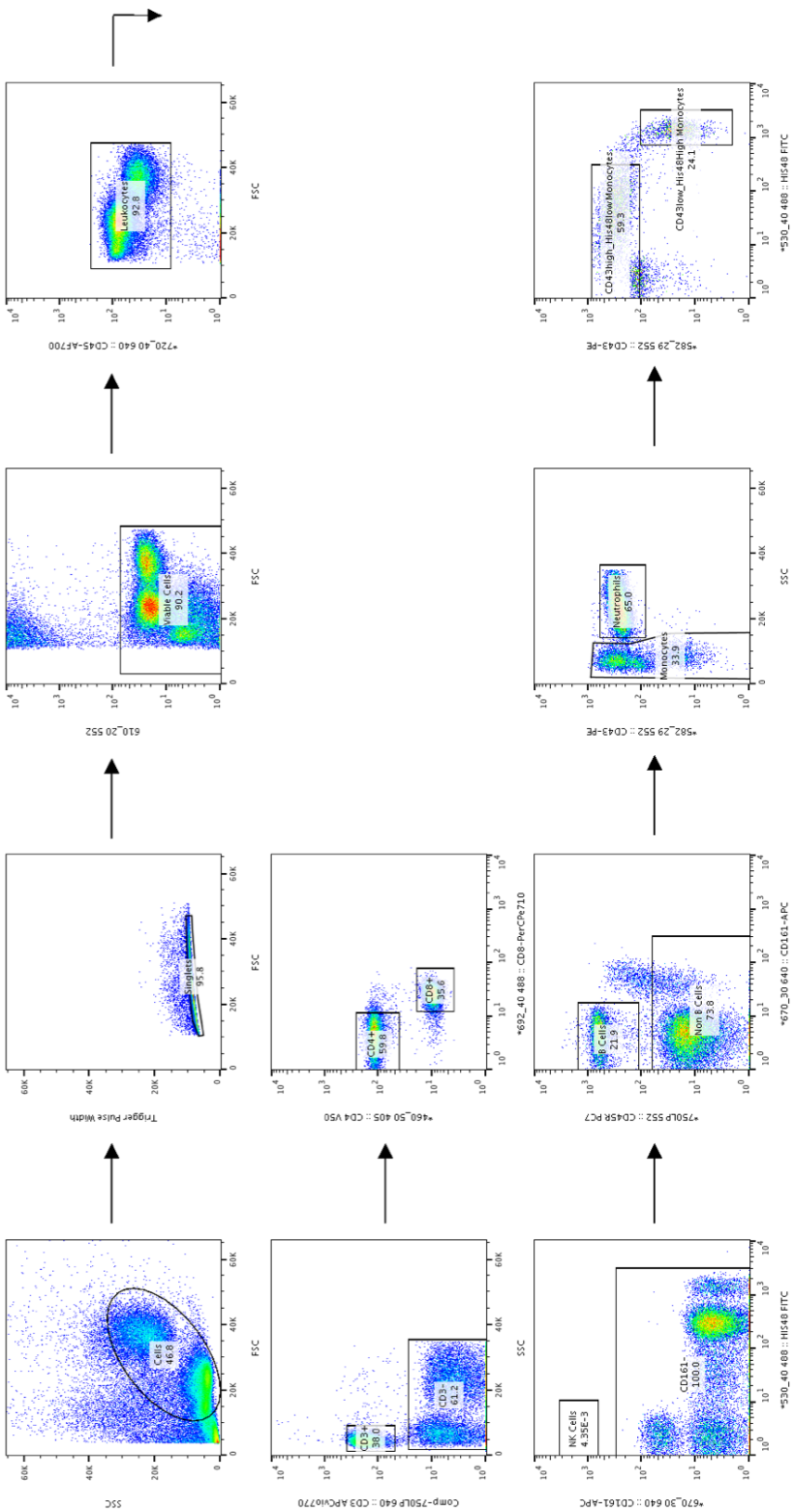


Figure 6.3: Representative gating strategy for 9-Colour FACS Analysis, as taken from Barnett-Vanes *et al.*[205] Initial gating compares levels of side-scatter (SSC) and forward-scatter (FSC) to identify cells from debris before singlet events are resolved by way of a low Trigger Pulse Width. Single cell events are then classified by viability, as detected by a low signal of the zombie staining method employed. Using the cell surface marker CD45, the leukocyte population can be isolated. The presence of the surface marker CD3 allows the resolution of total T-cells, which can then be separated by the detection of CD4 or CD8 in order to discern Cytotoxic T-cells from T-Helper Cells respectively. From the CD3- population, NK cells can be resolved by the presence of CD161. All remaining cells can be split into B-Cells and non B-cells by the presence of CD45R. A higher level of SSC is indicative of increased internal complexity (i.e. granularity) and allows for the non B-cells to be split into monocytes and neutrophils. Finally, the ratio of His48 and CD43 cell surface markers can be used to separate Classical from Non-classical monocytes.

6.4). A micropunch at 1mm diameter (Fine Science Tools) was used to isolate PVN samples in unstained sections and deposit them into a 1.5ml centrifuge tube ready for RNA extraction.

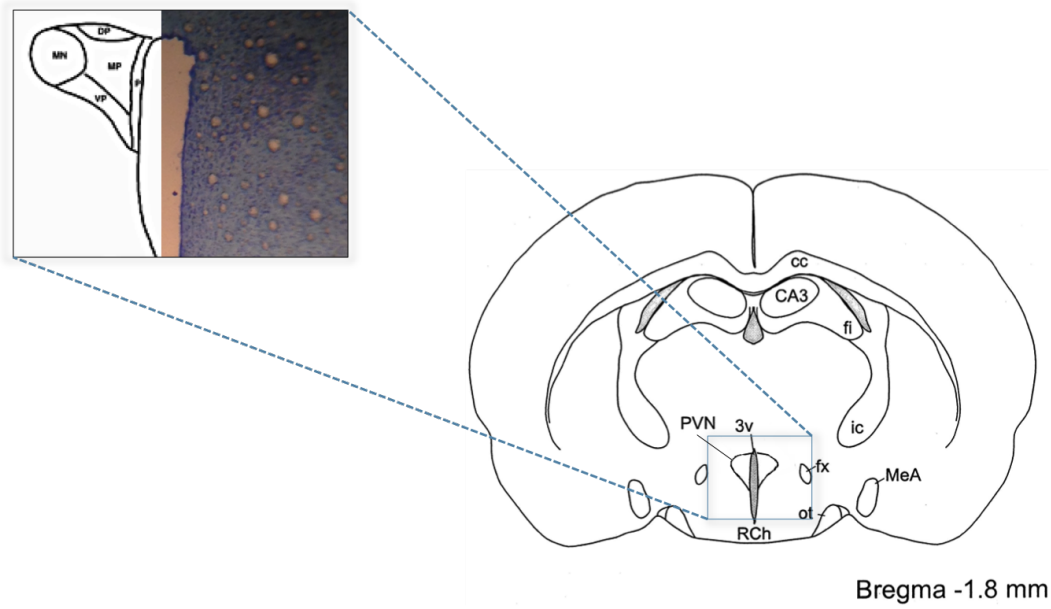


Figure 6.4: Coronal section of Rat Brain illustrating atlas coordinates and Toluidine Blue stain of the Paraventricular Nucleus (PVN). Here is a coronal section of the rat brain at Bregma -1.8mm showing the location of PVN lateral to the third ventrical. The PVN has been enlarged in the left side of the image to illustrate both a Toluidine Blue stain of the region (Right), and a depiction of how the nuclei are separated (Left; dorsal parvocellular subdivision (DP), magnocellular division (MN), medial parvocellular subdivision (MP), periventricular parvocellular subdividion (P), ventral parvocellular subdivision (VP).)

6.3.2 RNA Extraction

Similarly to the blood RNA extraction protocol, all procedures for PVN RNA extraction were conducted in an RNAse free environment ensured by good laboratory practice in addition to liberal cleaning of equipment and workspaces with RNAseZap (Ambion). Sample punches were homogenized in 1ml TRIzol reagent (Invitrogen), before centrifuging at $\sim$ 12,000 x g for 12 minutes at 4°C. Supernatants from each sample were carefully decanted into new tubes, leaving behind a pellet of cellular debris to be discarded.

One volume of EtOH (100% v/v) was added directly to one volume of sample homogenate in TRIzol reagent before vortexing to ensure thorough mixing. Then the sample mixtures were added into Zymo-Spin IIC Columns (Zymo Research) placed in collection tubes and centrifuged at $\sim$ 12,000 x g for 1 minute. Spin columns were then transferred to new collection tubes and 400μl DirectZol RNA PreWash buffer was added to the column. Samples were again centrifuged for 1 minute with the flow through discarded. This step was repeated. Next, 700μl RNA Wash Buffer was added to the columns and samples were centrifuged for 1 minute. Flow-through was discarded and samples were centrifuged again for 2 minutes to ensure complete removal of the wash buffer. Columns were then transferred to an RNase-free tube and 50μl of DNase/RNase-free water was added to the spin column. Samples were centrifuged for 1 minute to elute the RNA. Final concentrations of RNA for biomarker validation were confirmed using spectrophotometry (Nanodrop 2000c, Thermo Scientific).

6.3.3 cDNA Synthesis

All cDNA synthesis steps were carried out as outlined in section 2.3.4 with the Quantitect RT Kit (Qiagen).

6.3.4 Quantitative PCR Analysis

qPCR Analysis of Ifit1 expression in PVN was carried out as outlined in section 2.3.13.

6.3.5 9-Colour Panel Antibody Validation

Creation of an Ifit1 expressing construct

Initially, RNA was pooled from the SHR juvenile blood samples in order to maximise the levels of Ifit1 transcripts contained. A cDNA synthesis was carried out using the Qiagen Quantitect kit, as outlined in section 2.3.4. However, the random hexamer

primers supplied with the kit were swapped for an oligo $d(T)_{20}$ primer in order to capture full length mRNA transcripts containing a complementary Polyadenylated tail. Once a population of cDNA was produced, PCR Primers were selected to amplify and amend the Ifit1 transcript for eventual ligation and transfection (Table 6.2). The 3' boundary of the primers were designed to be comparible in size to those used for ifit1 in table 2.4.

Ifit1 BamH1 (F)	5' – CGCGGATCCGCCACCATGGGAGAGAATGCTGGTGGTGA
Ifit1 Xho1 (R)	3' – CCGCTCGAGGCTGCATTCAAAATGCAGGGTTCAT
Annotation;	Start Codon
	Kozak Consensus Sequence
	Ifit1 Specific
	BamH1 Digestion Site
	Xho1 Digestion Site
	Cleavage Overhang

Table 6.2: Primers for PCA amplification of Ifit1 Gene. Also shown are the specific regions incorporated for later digestion in addition to a Kozak Consensus Sequence for increased transcription.

Using the oligo $d(T)_{20}$ Primer synthesised cDNA as a template, a PCR reaction master mix was made up to 50μl total (Table 6.3). The sample was then ran on a thermocycler to amplify the Ifit1 transcript and add on the necessary cleavage sites required for insertion into a vector (Table 6.4). The PCR product was next run on a 1% Agarose gel to inspect whether it had been successful. A solution of 0.3g Agarose (Sigma Aldrich) was dissolved in 30ml of Millipore water, heated in the microwave, along with 1μl of Ethidium Bromide (Sigma Aldrich). This was mixed thoroughly and allowed to cool to approximately 50°C. This mixture was then poured into a small gel rig (Sigma Aldrich) with a loading comb in place and allowed to polymerise. Once polymerised, the gel rig was filled with 1x TAE buffer to cover the gel. 10μl of Quick-Load Purple 1Kb DNA Ladder (New England Biolabs) was loaded into the first well and two mixtures of 10μl PCR product, 2μl 6x Loading Dye (New England Biolabs) and 1μl Ethidium Bromide (Sigma Aldrich) were added into the adjoining lanes. The gel was run at 125v until the loading dye was approx. 80% through the gel. Images were taken using the UV transilluminator (Azure Biosystems) to confirm whether a product was seen at all,

and which size the band corresponded to. A size band corresponding to the ifit1 gene CDS (1466bp) was indicative of a successful PCR amplification (Figure 6.5).

Component	Volume	Final Concentration
5x Phusion HF Buffer	10μl	1x
Primers (Table 6.2)	0.2μ Each	200μM
10mM dNTPs	1μl	0.5μM
Nuclease-free Water	36μl	
Phusion DNA Polymerase	0.5μl	1.0 units/50μl PCR
cDNA Template	2μl	

Table 6.3: Reaction master mix for Phusion DNA PCR

Step	Temp	Time
Initial Denaturation	98°C	30 Seconds
Amplification (35 Cycles)	98°C	1 Minute
	72°C	30 Seconds
Final Extension	72°C	5 Minutes
Hold	4°C	∞

Table 6.4: PCR Amplification Protocol

The two bands were carefully excised from the gel using and a scalpel blade and a UV transilluminator to view the bands. Care was taken to remove as much agarose gel as possible. The containing DNA was then isolated using the Gel Extraction kit (Qiagen). The excised gel bands were weighed and 3 volumes of Buffer QC were added to 1 volume of gel (100mg - 100μl). This solution was incubated at 50°C for 10 minutes and vortexed every 2 minutes to ensure the agarose gel was full digested. As the adsorption of DNA to the QIAquick membrane is only efficient at acidic conditions of pH $\leq$ 7.5, the Buffer contains a pH indicator. A check of the solution's colour confirmed the pH to still be acidic, and so no action was required. The QIAquick spin column was then placed in a 2mL collection tube, and the solution was loaded into the column and centrifuged at $\sim$ 14,200 x g for 1 minute. Flow-through was discarded, before 500μl of Buffer QC was loaded directly onto the column and left for 5 minutes at room temperature. The column was then centrifuged again for 1 minute at $\sim$ 14,200 x g. Next, the column was washed with 75μl of Buffer PE and left for 3 minutes at room

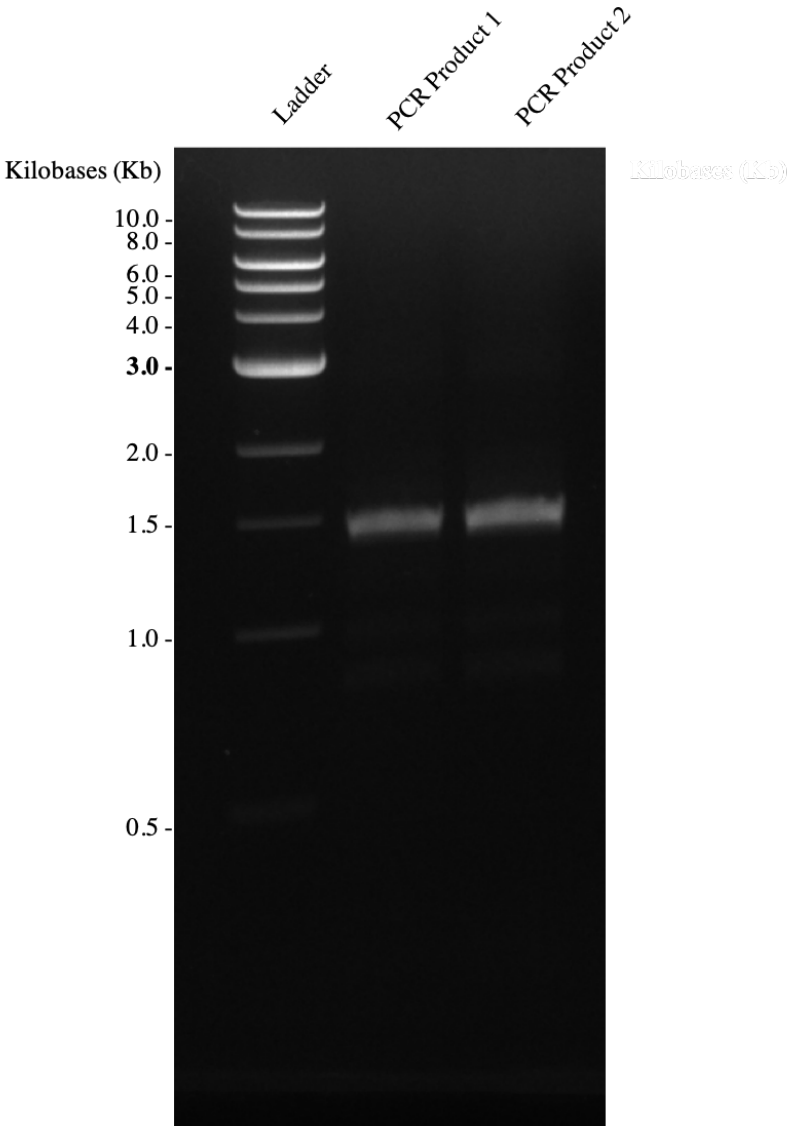


Figure 6.5: PCR Product of size matching that of Ifit1. Viewed on 1% Agarose Gel

temperature before centrifuging for 1 minute. Flow-through was again discarded, and the column was centrifuged once more for 1 minute to remove any residue. Finally, the column was placed into a clean 1.5mL microcentrifuge tube. To elute the DNA, 30μl of Buffer EB (10mM Tris-Cl, pH 8.5) was pipetted directly onto the centre of the membrane and left to stand for 1 minute. The sample was then centrifuged at $\sim$ 14,200 x g for 1 minute and DNA was eluted.

A tube of purified DNA and a pcDNA3.1(+) vector were incubated overnight at 37°C with a reaction buffer to produce the specific corresponding sticky ends required for ligation of the insert into the vector (Table 6.5). The digestion products were reran on a 1% agarose gel and excised, as outlined above, in order to purify the product. In order to optimise the ligation procedure, a concentration calculation was conducted. A 1% agarose gel was ran with Quick-Load Purple DNA Ladder (New England Biosciences) at varying loading concentrations (2μl, 4μl, & 8μl). The insert was loaded next at either 2μl or 4μl of volume (mixed with 3μl or 1μl of water respectively to maintain a loading volume of 5μl in total). The vector was loaded at either 1μl or 2μl of volume (made up to 5μl with water) in the next two loading wells. This gave the opportunity to make a comparison of concentration against the known concentrations of the maker lanes (Table 6.6).

Component	Volume
cDNA	28μl
XhoI	1μl
BamHI	1μl
Cut Buffer (10x)	5μl
Water	15μl

Table 6.5: PCR Digestion Mastermix

Using the concentrations inferred from the gel, a ligation molar ratio could be calculated of insert to vector using the equation outlined in Figure 6.7. A final quantity of 13.5ng was determined for a 1:1 ratio of Insert:Vector. Since the Promega website

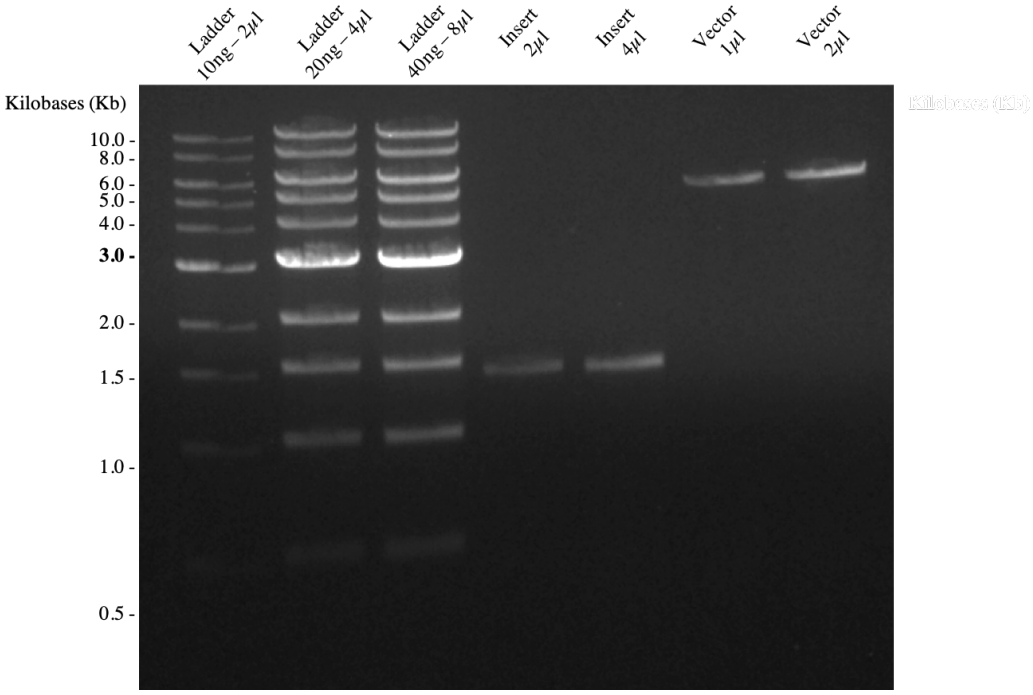


Figure 6.6: Semi-quantitative assessment of PCR inserts and vector

$$x = \frac{kb\ of\ insert}{kb\ of\ vector} \times ng\ of\ vector$$

Where; $Insert\ length = 1466bp$

$Vector\ amount = 50ng$, as defined by the T4 Ligase Protocol

$VectorLength = 5428$

Figure 6.7: Equation for calculating Ligation Molar Ratio

state ligations are tested at ratios from 1:1 to 3:1, and given the collective experience of colleagues within the laboratory, a ratio of 3:1 (40.5ng of insert) was selected to ensure an excess of insert without requiring more than 20μl of reaction volume as outlined in Table 6.6. This reaction mixture was made up in a microcentrifuge tube on ice and mixed gently by pipetting up and down. It was then incubated at room temperature (21°C) for 15 minutes before being placed back on ice to chill the reaction mixture.

Component	Volume for 20μl Reaction
T4 DNA Ligase Buffer (10x)	2μl
Vector DNA	50ng (2.5μl)
Insert DNA	40.5ng (10.8μl)
Nuclease free water	to 20μl
T4 DNA Ligase	1μl

Table 6.6: Ligation Mastermix for T4 DNA Ligase

Bacterial Transformation

The mixture containing ligated and unligated DNA could then be transformed into competent *Escherichia coli* (E. coli) DH5α cells (Thermofisher). 50μl of DH5α in calcium chloride suspension were thawed on ice and mixed with 5μl of the ligated plasmid. This mixture was incubated on ice for 50 minutes before being placed in a 42°C water bath for 45 seconds to heat shock the bacteria. This was then placed back on ice for 2 minutes. 450μl of Super Optimal Broth with glucose (SOC) medium was then added in order to dilute the calcium chloride, stabilising the cells. This was then left in a shaking incubator for 1 hour at 37°C. 100μl of the cells were plated onto an agar plate containing ampicillin and incubated at 37°C overnight. Using a pipette tip, three colonies were isolated and grown overnight in a 37°C shaking incubator in 5ml of Lysogeny Broth (LB) containing ampicillin. The resulting bacterial suspensions were passed through a QIAprep® Spin Miniprep Kit (Qiagen) to isolate plasmid DNA. 1ml of bacterial culture was pelleted by centrifugation at >6,800 x g for 3 minutes at room temperature (15-25°C). The pellets were next resuspended in 250μl of Buffer P1 and

transferred to a microcentrifuge tube. 250μl of Buffer P2 was then added and the mixture was inverted thoroughly until the solution became clear. Once the solution was clear, 350μl of Buffer N3 was added and the mixtures were immediately mixed thoroughly before centrifuging for 10 minutes at 17,900 x g. The supernatant from the previous step was decanted via a pipette to a QIAprep spin column and centrifuged for 30-60 seconds. The flow-through was discarded. via an additional centrifugation, 0.75ml of Buffer PE was used to wash through the spin columns and again the flow-through was discarded. The columns were centrifuged at the above speed for 1 minute in order to remove residual wash buffer. Next, the columns were transferred to a clean 1.5ml microcentrifuge tube and 50μl of Buffer EM (10mM Tris·Cl) was added to the centre of the column. These were left to stand for 1 minute before centrifuging for 1 minute to elute the plasmid DNA. A nanodrop spectrophotometer was used to quantify the resulting plasmid DNA (Table 6.7).

Colony	Conc (ng/μl)	A260	A280	260:230	260:230
A	205.7	4.11	2.23	1.84	1.83
B	276.4	5.53	2.99	1.85	1.78
C	347.5	6.95	3.72	1.87	1.97

Table 6.7: Spectrophotometry output for Isolated Ifit1 constructs

As the pcDNA3.1(+) vector confers ampicillin resistance, growth of colonies on the ampicillin containing agar only confirmed successful transformation of the plasmid. These may contain the insert of interest, or may well be re-ligated vectors. In order to confirm a successful ligation of the vector and insert, the plasmid DNA was initially incubated with BamHI and XhoI and ran on a 1% agarose gel to assess the size of bands produced from the digestion (Table 6.8, Figure 6.8). Three distinct bands are apparent from the BamHI and XhoI digestion at; 5,500bp, 3,500bp and 1,400bp. The whole pcDNA3.1(+) is 5428bp in size, which seems consistent with the strongest band at 5,500bp. This band appears to be either undigested, or simply does not contain the insert and so digestion only serves to linearise the plasmid as the BamHI and XhoI

sites are only 56bp away from each other. The band at 1,500 appears to be the Ifit1 insert itself, as the insert is 1466bp and directly flanked by BamHI and XhoI digestion sites.

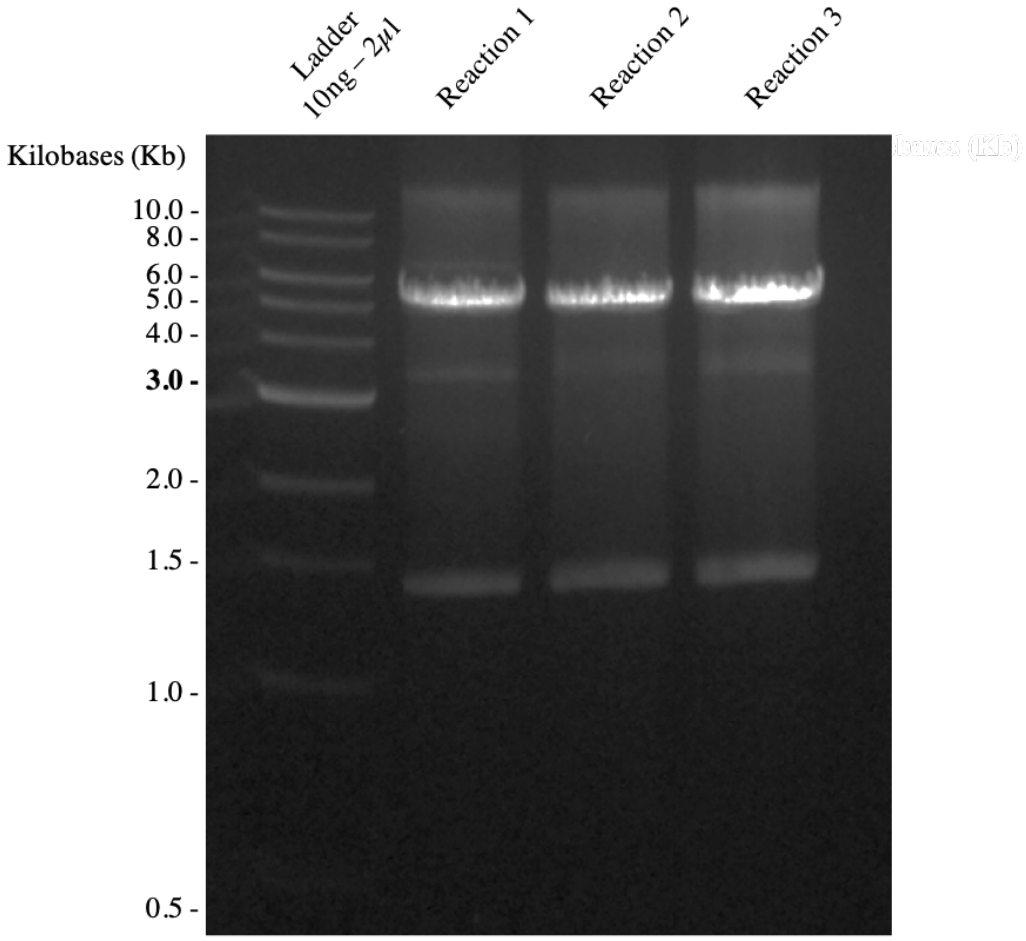


Figure 6.8: BamHI & XhoI Digest on Plasmid Construct

For additional confirmation, 1500ng of each sample DNA was sent for Sanger Sequencing with primers corresponding to either side of the ligation sites (Table 6.9). The results were curated and both ends of the sequencing were merged to give a final sequence of the inserts. These were then ran through a BLAST search algorithm to confirm the sequence identity with the NCBI Ifit1 Rattus norvegicus sequence. Each of the inserts showed a high sequence identity with Ifit1 sequence entry and so were deemed suitable for usage downstream (Appendix C).

Component	Volume per Reaction (μl)
BamHI	0.5
XhoI	0.5
Plasmid cDNA	5
Cut Buffer (10x)	5
Water	3

Table 6.8: Digestion Mastermix for confirming insert in plasmid construct

Primer Site	Sequence
T7 Promoter	*5'-TAATACGACTCACTATAGGG-3'*
pCR3.1-BGHrev	*5'-TAGAAGGCACAGTCGAGG-3'*

Table 6.9: Sanger Sequencing of Plasmid Primer Sequences

Transfection into HEK293t Cell line

Six wells of a 24-well plate were seeded with carefully thawed Human Embryonic Kidney Cells, expressing a mutant version of the SV40 large T antigen (HEK293T). They were maintained in Dulbecco's Modified Eagle Medium containing pH indicator, with additional L-Glutamine (4mM), Streptomycin (1%), and Fetal Bovine Serum (FBS, 10% Thermofisher). These were maintained with fresh medium and incubated at 37°C until they reached approximately 70% confluency. At this point, a lipofectamine transfection could be carried out of either control plasmid (pcDNA3.1(+)) or the plasmid construct containing rat Ifit1 (pcDNA3.1(+)·Ifit1) as per the manufacturer's instructions (Invitrogen).[206] In two 1.5ml microcentrifuge tube, 4.5μl Lipofectamine 3000 Reagent was diluted in 150μl Opti-MEM Reduced Serum Medium and mixed well. In two additional tubes, 150μl of Opti-MEM media was mixed with 1.5μg of respective DNA (1.5μl pcDNA3.1(+), 3.78μl pcDNA3.1(+)·Ifit1). To both of these, 3μl of p3000 Reagent was added. Each of the tubes were combined with one of the previously prepared Lipofectamine tubes and left to incubate at room temperature for 5 minutes. Serum containing media was decanted from each of the wells on the plate, and replaced with Opti-MEM Reduced serum medium before dividing each lipofectamine

mixture between three wells (3 x pcDNA3.1(+), 3 x pcDNA3.1(+)·Ifit1) and incubating at 37°C. Cells were checked several times daily to prevent toxically high confluence from occurring. After 2-days of expression, cells could be lysed to confirm protein expression.

Confirmation of Protein Expression

Transfected cells (HEK293T Control and HEK293T·Ifit1) were lysed on their respective cell culture dishes using 150μl of lysis solution consisting of sodium dodecyl sulfate (SDS), Tris·HCl and HEPES·NaOH. After the solution was applied to each of the plates, a cell scraper (Sigma-Aldrich) was used to detach the cells from the plate and ensure thorough lysis. Cell lysates were collected into a 1.5ml centrifuge tube, and passed through a 5mL syringe with 25G needle to break up the cells further.

Protein concentrations of both lysates was measured using the Nanodrop 2000c (Thermofisher). This was calculated by assessing A280 absorbance values with 1 Abs being equal to 1mg/mL. Protein concentrations were relatively equal and therefore did not warrant further dilutions to normalise between the two groups.

Next, 50μl of 4x loading buffer (Table 6.10) was added to each of the sample lysates to make loading of the samples into their respective wells easier, as well as enabling visualisation of the bands as they progress through the electrophoresis gel. Sample lysates were next heated on a heat block to 95°C briefly to ensure any SDS had not precipitated in the lysate mixes and the proteins were properly denatured prior to gel separation.

A 10% electrophoresis gel was prepared (Table 6.10), with a 4% stacking gel component. This was prepared along with 1L of running buffer.

Along with the HEK293T and HEK293T·Ifit1 sample lysates, lysates from two other cell lines were included to assess endogenous expression of ifit1 for potential involvement in further validation studies. These were the Att20 mouse pituitary cells and the PC12 rat adrenal medulla cells and were available due to ongoing experiments of colleagues within the laboratory. All four cell lysates were loaded at both 30μg and

60μg concentrations, as well as 10μl of protein size ladder (Novex™ Sharp Pre-Stained Protein Standard, Thermofisher).

The gel was run for approximately 45 minutes at 180v, until the size markers had separated sufficiently to easily resolve the 50kDa marker. At this point a semi-dry protein transfer was conducted using the Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad). A transfer buffer was initially prepared (Table 6.10). A PVDF membrane (Millipore) was first activated in 100% Methanol for 30 seconds before rinsing with dH2O and submerging in transfer buffer to acclimatise prior to the transfer. Four sheets of filter paper were used to sandwich the PVDF membrane and electrophoresis gel. The stack was soaked in transfer buffer and ran at 20v for 110 minutes. The PVDF membrane was soaked in Ponceau Red reagent (0.1% w/v Ponceau, 5% v/v Acetic acid) for approximately 10 minutes then rinsed in dH_2O to visualise the protein transfer was successful. The membrane was destained with several washes in bovine serum albumin (BSA) in phosphate buffered saline (PBS) (5% w/v) ready for antibody visualisation (Figure 6.9).

Immunocytochemistry

A newly transfected population of cells (as outlined above) were grown on glass cover slips (Thermanox - Nalge Nunc International, NY, USA). Growth media was aspirated from them before being gently washed with 1mL of sterile PBS. This was carefully removed, and 1mL 4% Paraformaldehyde (PFA) was added to fix the cells. This was left on for 15 minutes at room temperature. Following fixation, cells were washed a further two times with fresh PBS. Cells could then be blocked and permeabilised by incubation for 10 minutes with permeabilisation buffer (2% Bovine Serum Albumin (BSA), 0.1% Triton X-100). Primary Anti-Ifit1 (OTI3G8, Novus Biologicals) diluted in PBS containing 5% BSA was dropped onto parafilm strips in a light-sealed box. Coverslips could then be removed from the cell culture plates and inverted ontop of the parafilm to enable an economical usage of primary antibody. Coverslips were light-sealed

Buffer	Constituents	Quantity
2x Separation Buffer	Tris Base	9.08g
	SDS	0.2g
	HCl/NaOH	pH to 8.8
	H_2O	Up to 100mL
2x Stacking Buffer	Tris Base	3g
	SDS	0.2g
	HCl/NaOH	pH to 6.8
	H_2O	Up to 100mL
1x Separation Gel (10%)	2x Separation Buffer	5mL
	BisAcrylamide (40%)	2.5mL
	H_2O	2.5mL
	TEMED	20μL
	APS (10% in H_2O)	100μL
1x Stacking Gel (4%)	2x Stacking Buffer	1mL
	BisAcrylamide (40%)	170μL
	H_2O	824μL
	TEMED	4μL
	APS (10% in H_2O)	20μL
10x Running Buffer	Tris Base	30.3g
	Glycine	144g
	SDS	10g
	HCl/NaOH	pH to 8.3
	H_2O	Up to 1L
1x Transfer Buffer	Tris Base	3g
	Glycine	15g
	Methanol	200mL
	HCl/NaOH	pH to 8.3
	H_2O	Up to 1L
4x Loading Buffer	1M Tris HCl	2mL
	SDS	0.8g
	Glycerol	4mL
	0.5M EDTA	1mL
	Bromophenol Blue	8mg
	2M Dithiothreitol	2mL
	HCl/NaOH	pH to 6.8

Table 6.10: Reagents used for Western Blot analysis of Ifit1 protein expression

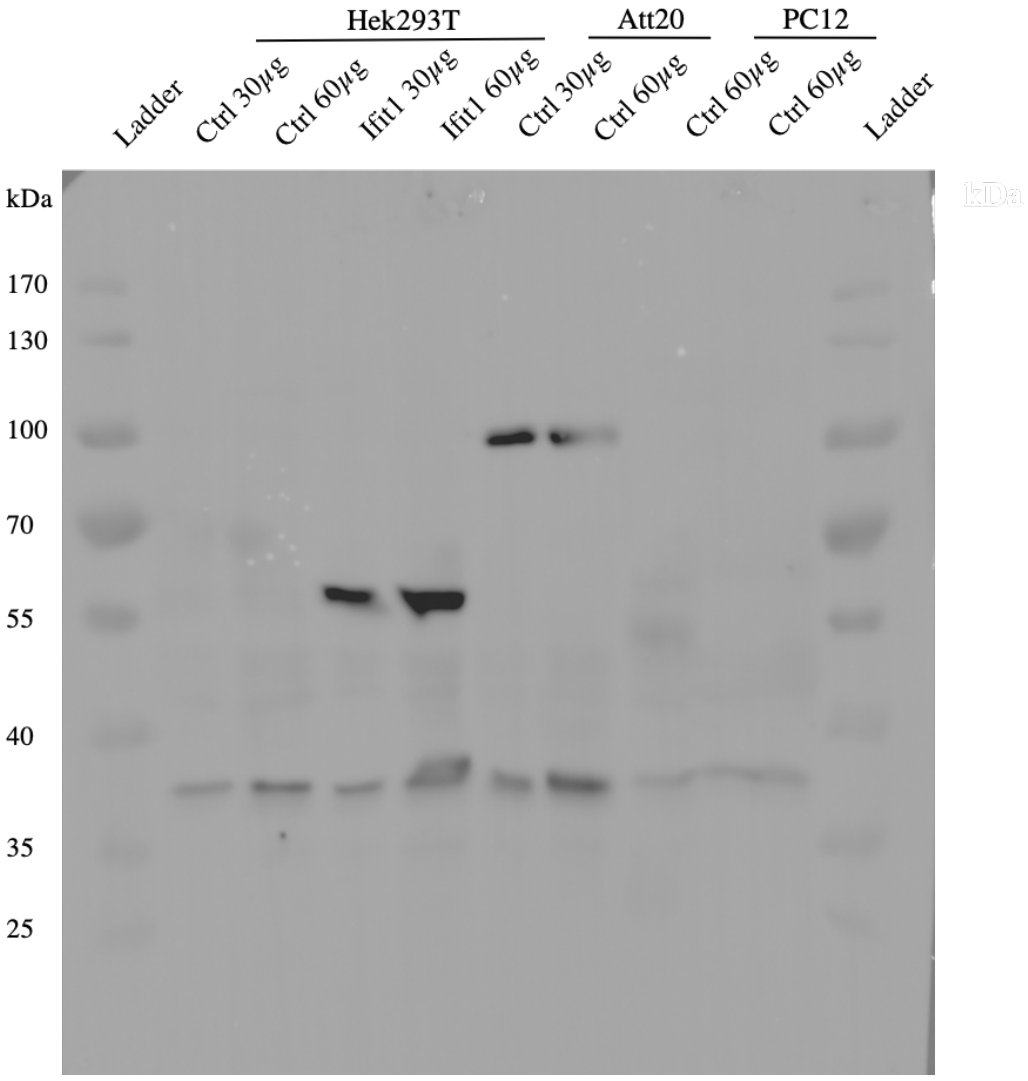


Figure 6.9: Antibody test on various cell lysates. Here Hek293t cells, either transfected with Ifit1 or a control pcDNA3.1 construct, were lysed and ran on a Western Blot before being incubated with the Anti-Ifit1 antibody. Also featured are cell lysates from Att20, and PC12 cells. Two clear bands can be seen just above the 55kDa size, a size consistent with the Ifit1 protein. There also appears to be a distinct band in both of the Att20 lysates at approximately 100kDa in size. This may be a product of ifit protein dimerisation, due to insufficient SDS denaturation. Indeed the IFIT2 protein has been shown to dimerise *in vivo*.[207] However the evidence for IFIT1 dimerisation is less clear. Further characterisation of these bands would be possible through extraction and interrogation using Mass Spectrometry techniques.

and left overnight at 4°C before being washed with PBS three times. Using the same inverted incubation method coverslips were incubated with secondary antibodies diluted in PBS with 5% BSA for 1 hour at room temperature in the dark. After this incubation period DAPI was added and samples were left for a further 10 minutes. Following three additional washes in BSA samples were imaged on a DMI6000B microscope (Figure 6.10; Leica - Wetzlar, Germany).

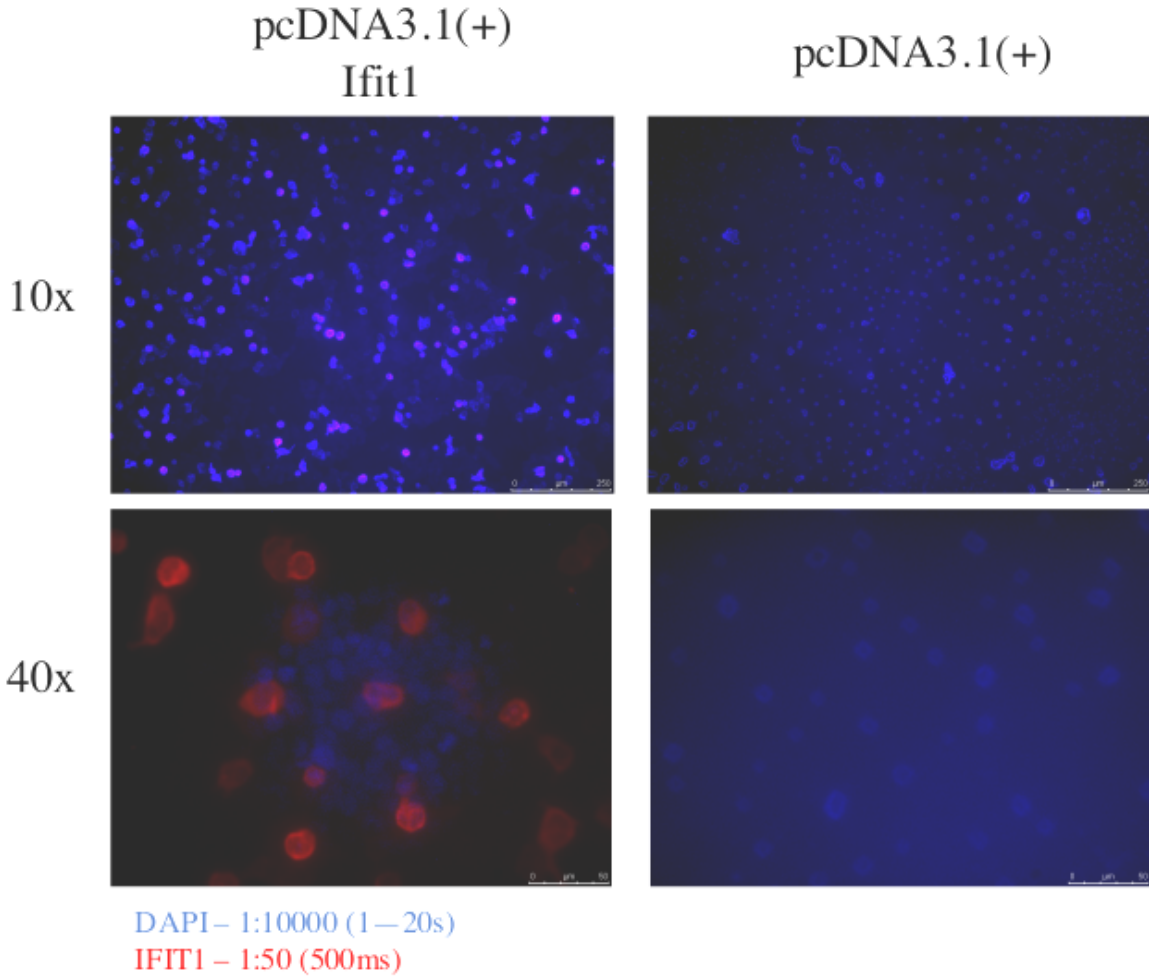


Figure 6.10: Antibody test on HEK293T cells containing either control transfection (pcDNA3.1(+)) or Ifit1 transfection (pcDNA3.1(+)Ifit1). Cells were incubated with anti-Ifit1 antibody (1:50, 500ms exposure, red) and DAPI (1:1000, 1-20s exposure, blue) before being imaged at 10x and 40x magnification. Here the ifit1 transfected cells show a clear signal in the red channel whereas no signal can be seen in the control transfected cells.

6.3.6 Fluorescence-activated cell sorting

Antibody Panel Viability

Once antibodies had been delivered, the first step was to assess their integrity was not compromised during transit, and that their binding efficiency was suitable for my intended downstream application. To do this, 2μl of each antibody was added to a glass tube containing 1 drop of OneComp eBeads™ Compensation Beads (ThermoFisher Scientific). Each drop of beads contains two populations; a positive population that has reactivity with any rat, mouse or hamster antibody, and a negative population that has no reactivity towards any antibodies. Firstly, a solution of unstained beads was ran through the cytometer to show the FSC/SSC of 10,000 events. This allowed a single-cell gate to be drawn across the relevant proportion of events prioritised for antibody optimisation, preventing any cohered beads from passing through at the same time and flagging a false positive/negative. Then each antibody containing tube was ran individually through the FACS machine (BD Fortessa X20 analyser, BD Biosciences - Franklin Lakes, NJ USA) with their corresponding excitation wavelength laser and filter parameters set up to a measured total of 10,000 events (Table 6.11).

The resulting emission profiles showed a bimodal distribution for each antibody, with a clear separation observed between the negative and positive peaks. Care was taken to ensure each peak lay on a linear, and not logarithmic, measure apart. This keeps the final measurements within the dynamic range of the antibodies, allowing quantities to be measured more accurately than if binding was outside of the linear range. This process also enables a pragmatic test as to whether there is any cause for concern with the panel design by highlighting any potential crossover of emission bands. Based on the antibody emissions from this optimisation, the software can mitigate this by creating compensation parameters to better resolve peaks that show similar profiles.

Antibody	Indicates	Fluorochrome	Supplier	Clone	Suggested Dilution	Optimised Dilution	Excitation	Emission
Live-Dead	Cell Viability	APC-Cy7	Biolegend	NIR Zombie	1:1000	1:1000	633nm	746nm
His48	Neutrophil	FITC	BD Pharminogen	HIS48	N/A	1:400	488nm	520nm
CD43	Monocytes	PE	Biolegend	W3/13	1:100	1:1000	496nm	578nm
CD161	NK Cells	APC	Biolegend	3.2.3	1:1000	1:200	633nm	650nm
CD4	T-Helper Cells	V450	BD Biosciences	OX-35	1:10-50	1:100	404nm	448nm
CD8a	T-Cytotoxic Cells	PerCP-eF710	Biolegend	OX-8	1:20	1:75	482nm	675nm
CD3	T-Cells (All)	APC Vio770	Miltenyi Biotec	REA223	1:10	1:10	652nm	776nm
CD45R	B-Cells	PC7	eBioscience	HIS24	N/A	1:160	488-561nm	775nm
CD32	FcγII	N/A	BD Pharminogen	D34-485	1:100	1:50	N/A	N/A
CD45	B-Cells	AlexaFluor 700	Biolegend	OX-1		1:50	696nm	719nm

Table 6.11: Panel design for FACS isolation of various Leukocyte populations in rat whole blood

Tissue Isolation

Following a week habituation period, 3 WKY and 3 SHRs at 12 weeks old were sacrificed by a cranium strike stun and guillotine. Whole blood was collected into sodium citrate containing tubes (BD Vacutainer) on ice to prevent coagulation and RNA degradation respectively. To obtain the plasma layer, blood was transferred into 2mL microcentrifuge tubes and spun at 1300 x g for 10 minutes at 4°C. The upper plasma layer was removed and frozen at -80°C for further experiments. The next step was to perform a red blood cell (RBC) lysis preparation to deplete the samples from erythrocytes that could interfere with the dynamic range of the FACS experiment. This was carried out using an ammonium chloride based Lysing Buffer (BD Biosciences) as per the manufacturer's recommended protocol. For every 200μl of whole blood, 2.0mL of lysing solution was added. Tubes were gently vortexed immediately after adding the lysing solution, and incubated at room temperature for 15 minutes in an area out of the light. Once samples were no longer cloudy, they were centrifuged for 5 minutes at 200 x g. The supernatant was carefully aspired without disturbing the pellet, and discarded. FACS buffer was made up for diluting antibodies and consisted of 0.5% BSA (Sigma-Aldrich) in PBS (Sigma-Aldrich). This solution was passed through Millipore filters in order to sterilise it. The now RBC depleted pellet was resuspended in FACS buffer for flow cytometric analysis.

In order to assess both the success of the RBC depletion, and calculate how many leukocytes are present in the final FACS suspension, hemocytometry was conducted. The population of leukocytes was obtained, suspended in FACS buffer and integrity was verified, antibodies could be added to aid in identification and separation of the various cellular factions present. This was initially done by taking a 100μl aliquot of the undiluted RBC depleted FACS suspension and mixing with 100μl of Trypan Blue Solution (0.4%, Sigma) for identification of non-viable cells that have ruptured. A cover slip was placed on top of the Bright-Line™ hemocytometer (Sigma-Aldrich)

and aligned so it rested balanced on the two spacing risers either side of the count matrices. 50-100μl of the FACS suspension/Trypan Blue mix was pipetted onto both of the counting matrices to create a drop of the stained suspension on each. Using a light microscope at 100x magnification (Nikon Eclipse TS100) and a hand counter, I was able to count cells within each of the five large 5x5 grids. Cells were excluded if they were touching the upper or left boundaries of the matrix and care was taken to discern between trypan blue stained cells (ruptured) and non-stained cells in the final counts. Then using the below formula, a calculation of viable cells per mL was made (Figure 6.11).

$$Cell\ Viability = \frac{Number\ of\ Live\ Cells}{Number\ of\ DeadCells}$$

$$Measured\,Cell\ Density\ (Cells/mL) = \frac{Average\ Cells\ per\ Square \times Dilution\ Factor}{Volume\ of\ a\ small\ square\ (mL)}$$

$$Total\ Viable\ Cells = Cell\ Density\ (Cells/mL) \times Volume\ (mL) \times Viability$$

$$Final\ Volume\ (mL) = \frac{Total\ Viable\ Cells}{Target\ Density\ (Cells/mL)}$$

Figure 6.11: Calculations for Cell Abundance and Viability

Optimisation of Antibody Dilutions

Before carrying the final FACS sorting experiment, it was pertinent to optimise the antibody dilutions so as to ensure sufficient antibody is present and to minimise non-specific binding by excess probing. To do this, one of the samples obtained from the SHR blood was selected for an antibody titration. A serial dilution of each antibody in FACS buffer was first made with the recommended dilution being used as the starting concentration. The dilutions tested are listed in Table 6.11.

Between 200,000-500,000 cells used for each antibody dilution. Cells were pelleted by spinning at 200g for 5mins and supernatant was removed, whilst being careful not to dislodge the cell pellet. 25μl of Anti-CD32 antibody was added at a concentration of 1:50 to the cell pellet and mixed. This was done in order to block antibodies binding non-specifically to the FcγII receptor. The solution was incubated for 10 minutes at room temperature. Following this incubation, 25μl of each antibody dilution (in FACs buffer) was added to each respective tube containing FcγII-blocked cells. These were briefly mixed by vortexing before incubated on ice, and in the dark, for 30 minutes. Cells were washed with 500μl of FACs buffer and vortexed before re-spinning to once again pellet the cells. Supernatants were discarded carefully, and cells were resuspended in 300μl of FACs buffer ready for cytometry assessment of antibody concentrations.

The suitability of each antibody dilution was mathematically assessed by use of the stain index (Figure 6.12, Table 6.12). The highest stain index denotes the concentration most able to discern between an antigen positive and negative population. For this reason, efforts were made to pick dilutions that would give a suitably high stain index, whilst ensuring an economical usage of the antibody. The antibody dilutions selected are highlighted in Table 6.12 in bold.

$$Stain\ Index = \frac{[MFI_1 - MFI_2]}{2\sigma}$$

Where; MFI = Mean Fluorescence Intensity
σ = Standard Deviation of Negative MFI

Figure 6.12: Equation to calculate the Stain Index for Fluorescently conjugated Antibodies

Antibody	Dilution	Mean Fluorescence Intensity (Positive)	Mean Fluorescence Intensity (Negative)	Mean Fluorescence Intensity (Negative) SD	Staining Index
CD4	**1:100**	**2476**	**121**	**78.3**	**15.04**
	1:200	1579	113	77.2	9.49
	1:400	998	139	96.6	4.45
	1:800	682	148	89.2	2.99
	1:1600	1034	173	104	4.14
CD45	**1:50**	**3448**	**55.2**	**126**	**13.46**
	1:100	**2442**	**62.9**	**127**	**9.37**
	1:200	1573	57.8	126	6.01
	1:400	946	62.9	116	3.81
	1:800	642	199	144	1.54
CD3	**1:10**	**500**	**8.98**	**21.4**	**11.47**
	1:20	453	8.98	20.8	10.67
	1:40	230	6.42	19.2	5.82
	1:80	127	5.13	17	3.58
CD8a	1:300	8318	285	537	7.48
	1:600	**8524**	**275**	**541**	**7.62**
	1:1200	2136	245	484	1.95
	1:2400	1835	266	460	1.71
CD43	**1:1000**	**4167**	**166**	**179**	**11.18**
	1:2000	1998	146	131	7.07
	1:3000	1751	139	122	6.61
CD45R	5:200	13144	136	162	40.15
	5:400	7629	88.7	118	31.95
	5:800	**3769**	**53.9**	**82.6**	**22.49**
	5:1600	2253	38.5	67.8	16.33
	5:3200	1012	29.5	53.2	9.23
CD161	1:100	1784	37.2	66.9	13.06
	1:200	**1748**	**28.2**	**59.1**	**14.55**
	1:400	1680	23.1	54.1	15.31
	1:800	1558	21.8	53.8	14.28
	1:1600	1174	21.8	51.8	11.12
His48	1:100	4623	46.2	54.2	42.22
	1:200	7529	43.6	54	69.31
	1:400	**6896**	**43.6**	**51.3**	**66.79**
	1:800	4416	43.6	53.4	40.94
	1:1600	2707	44.9	55.9	23.81

Table 6.12: Optimisation of Antibody Dilutions via the calculation of a Stain Index for each

6.3.7 FACS Analysis on Permeabilised Cells

Validation of Biotinylated Anti-Ifit1 Antibody

In order to attempt a positive selection of Ifit1 expressing blood cells, and to ensure it could be incorporated into the 9-panel FACS protocol, a biotinylated form of the same anti-Ifit1 clone was employed. To validate its suitability in a FACS protocol, the Control (pcDNA3.1) and Ifit1 (pcDNA3.1(+)·Ifit1) transfected cells were obtained via the above protocol. Initially, FACS buffer (0.5% BSA in PBS), Fixation buffer (4% PFA in PBS), and permeabilisation buffer (0.1% Triton X-100, 0.5% BSA in PBS) were all made up using autoclaved PBS solution.

After 48 hours since transfection, cells were Trypsin-detached (0.25%, sterile-filtered, BioReagent 2.5 g porcine trypsin and 0.2 g EDTA, Sigma-Aldrich) and pelleted at 180 x g for 3 minutes. These populations were washed with 1mL of sterile PBS and pelleted again. This wash step was repeated. Cells were counted using a hematocytometer as outlined above and resuspended at approx. $1x10^6$ Cells/100μl PBS. At this point the cells could be incubated with 1μl of Zombie dye per 100μl of cells for 15-30 minutes at room temperature in the dark. They were then washed with 2mL of FACS Buffer and resuspended to approx $3.24x10^6$ Cells/mL in ice cold FACS buffer. The following fixation method was adapted from Abcam's FACS Protocol database and conducted in the dark to preserve fluorescence of the Zombie Stain.[208] Cells were pelleted at 100 x g for 2 minutes before being resuspended in Fixation Buffer for 10-15 minutes. Cells were then washed with 2mL of Permeabilisation Buffer for 15 minutes and centrifuged at 100 x g for 5 minutes. The supernatant was discarded and they were resuspended in 2mL of 1% BSA in PBS. The Ifit1 transfected cells were divided into 5 groups for the assessing the following antibody dilutions; 1:50, 1:100 (Recommended), 1:200, 1:400, 1:800 (Figure 6.13). The control transfected cells were divided into 10 groups, as a no-primary antibody control was to be used for each dilution. Primary biotinylated anti-Ifit1 antibody was added to each of the populations at the above dilutions, apart

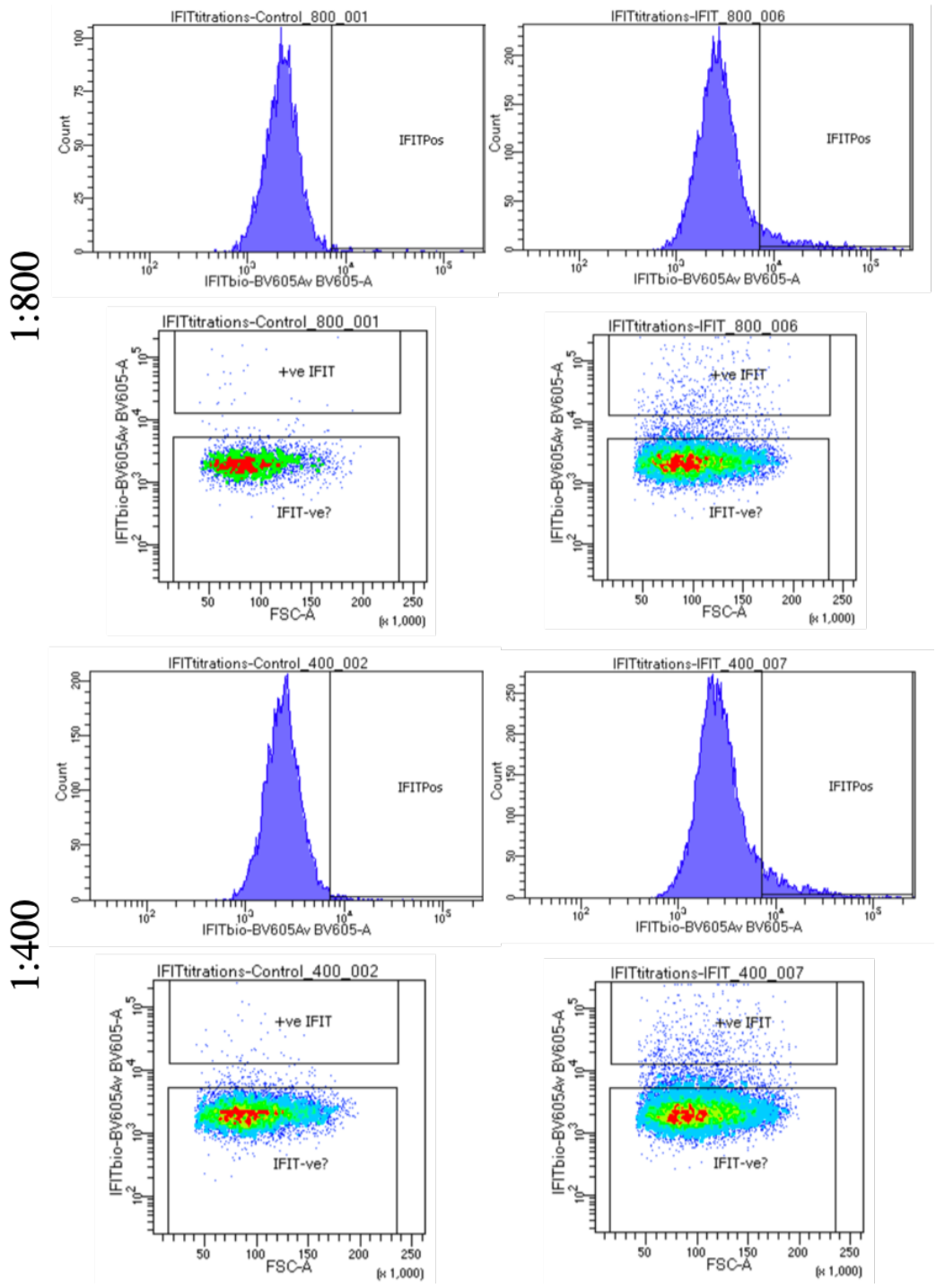


Figure 6.13: FACS Validation of Dilutions for Biotinylated Anti-Ifit1 Antibody

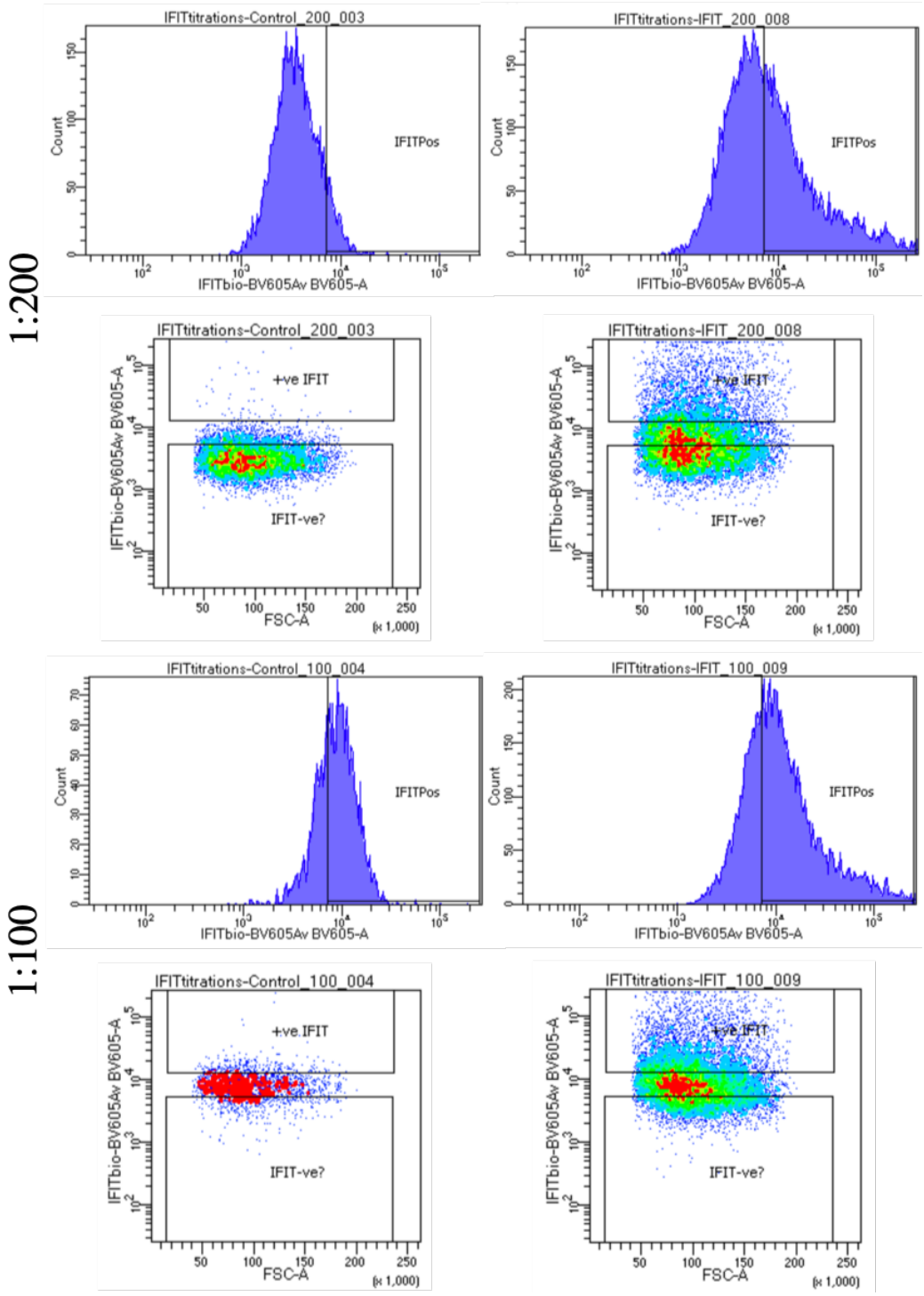


Figure 6.13: FACS Validation of Dilutions for Biotinylated Anti-Ifit1 Antibody

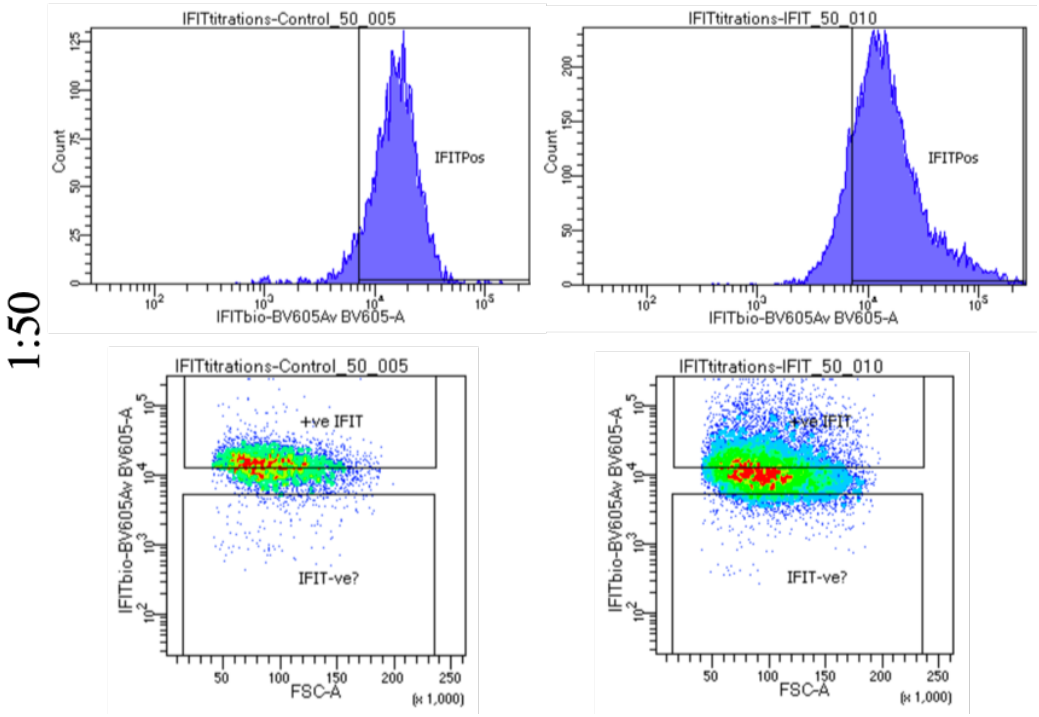


Figure 6.13: FACS Validation of Dilutions for Biotinylated Anti-Ifit1 Antibody

from the no-primary controls. The groups were left to incubate for 1 hour at room temperature. Samples were then washed with incubation buffer (1% BSA in PBS) and centrifuged at 100 x g for 3 minutes to pellet. The supernatant was carefully aspirated and discarded, and the wash step was repeated. Cell pellets were then resuspended in 100μl streptadvidin conjugated secondary, diluted 1:400 in 0.5% BSA containing PBS (BV605 Streptavidin, BD Horizon). These were then left for 30 minutes incubating at room temperature in the dark. Samples were washed twice in incubation buffer as outlined above before being stored at 4°C in 0.5% BSA PBS with 1mM EDTA and 0.1% sodium azide ready for FACS analysis.

6.4 Results

6.4.1 Ifit1 Expression in the PVN

Figure 6.14 shows the relative expression of Ifit1 in punched PVN sections between the WKY and SHR, as both juvenile and adult ages. Similarly to within blood tissue, Ifit1 shows a marked increase in abundance in the SHR as compared with the WKY, across both ages tested (Figure 6.14; juvenile WKY vs SHR, +10.1 fold, *P-Value<0.0001*; adult WKY vs SHR, +15.5 fold, *P-Value=0.0002*).

6.4.2 FACS Analysis to Isolate Cell Types

Novel distribution of cell types across strains

The initial sorting protocol revealed novel differences in Leukocyte abundance between the WKY and SHR bloods (Figure 6.15; Table 6.13).

Cells expressing CD3+ appeared subtly more abundant within the WKY blood than within the SHR blood, however this was by a small degree and did not reach significance (Figure 6.13; WKY vs SHR, -0.21 fold, *P-Value=0.254*).

The CD4+ marker was again subtly different within the WKY blood than in the SHR blood, but this appears to be due to random spread (Figure 6.13; WKY vs SHR, -0.14 fold, *P-Value=0.451*).

There was a significant downregulation of CD8+ cells in the SHR blood, with a fold change of -0.36 (Figure 6.13; WKY vs SHR, -0.32 fold, *P-Value=0.039*).

B-Cell populations did not appear to vary between the two strains, and remained relatively constant at around 16% of the total leukocyte population (Figure 6.13; WKY vs SHR, +0.04 fold, *P-Value=0.704*).

Interestingly, total monocytes showed a trend towards increased abundance within the SHR. Whilst this was not significant, the breakdown of all monocytes into CD43Hi/His48Lo and CD43Lo/His48Hi subpopulations revealed the CD43Hi/His48Lo monocytes to be

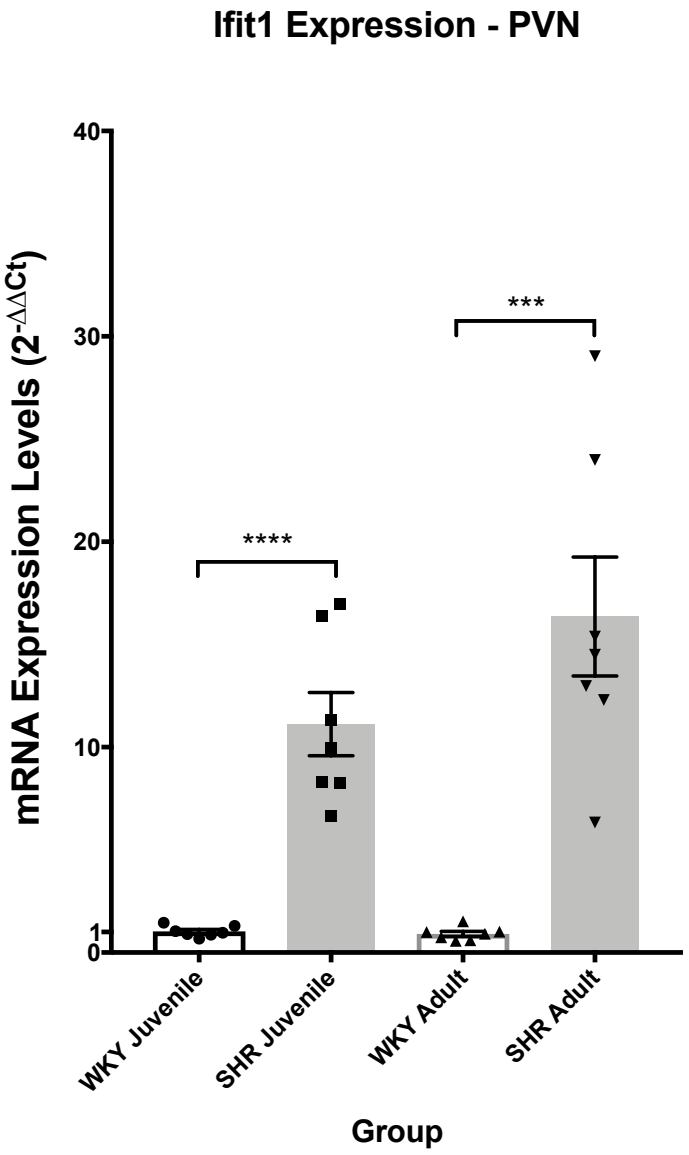


Figure 6.14: qPCR Relative Expression of Ifit1 in the PVN, displayed as $2^{-\Delta\Delta C_T}$ values. Statistical analysis made use of a Student's T-Test (P-Value <0.05, *; <0.01, **; <0.001, ***; <0.0001, ****).

significantly upregulated in the SHR (Figure 6.13; WKY vs SHR, Monocytes +0.61 fold, *P-Value=0.053*; CD43Hi/His48Lo +0.96 fold, *P-Value=0.014*; CD43Lo/His48Hi +0.37, *P-Value=0.287*).

The FACS gating strategy for Neutrophils showed very little variation between the two strains, with both having neutrophils accounting for approx. 18% of the total leukocyte population (Figure 6.13; WKY vs SHR, -0.14 fold, *P-Value=0.670*).

Finally, the strongest change in abundance was observed with the NK Cells as the WKY expressed negligible levels whereas NK Cells accounted for approx. 3% of the total leukocyte population in the SHR (Figure 6.13; WKY vs SHR, *P-Value<0.001*).

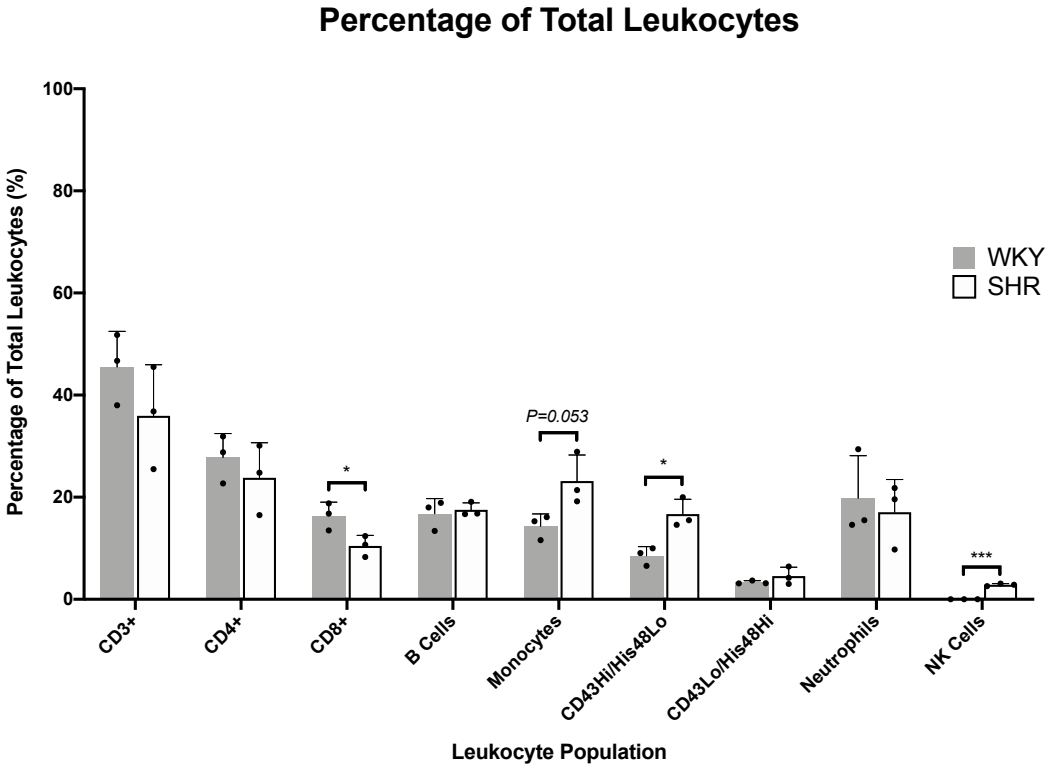


Figure 6.15: Differences in leukocyte populations across WKY and SHR strains. Statistical analysis made use of a Student's T-Test (P-Value <0.05, *; <0.01, **; <0.001, ***; <0.0001, ****).

	Percentage of Total Leukocytes								
	CD3+	CD4+	CD8+	B Cells	Monocytes	CD43Hi/His48Lo	CD43Lo/His48Hi	Neutrophils	NK Cells
SHR	25.50%	16.50%	8.28%	16.70%	28.90%	20.00%	6.42%	21.80%	3.09%
	36.80%	24.80%	10.70%	16.80%	21.40%	15.50%	4.21%	19.60%	2.58%
	45.50%	30.10%	12.40%	19.10%	19.20%	14.60%	3.03%	9.76%	2.84%
Mean	35.93%	23.80%	10.46%	17.53%	23.17%	16.70%	4.55%	17.05%	2.84%
WKY	51.80%	31.90%	18.80%	18.90%	11.60%	6.54%	3.11%	14.60%	0.00%
	46.70%	28.80%	16.80%	18.00%	16.10%	9.98%	3.14%	15.50%	0.00%
	38.00%	22.70%	13.50%	13.40%	15.30%	9.07%	3.69%	29.40%	2.66E-05%
Mean	45.50%	27.80%	16.37%	16.77%	14.33%	8.53%	3.31%	19.83%	8.87E-06%
P-Value (T-Test)	0.254	0.451	0.039	0.704	0.053	0.014	0.287	0.670	<0.001

Table 6.13: Statistical Comparison of Leukocyte Populations in WKY and SHR blood. Shown here are the relative abundances of cell types, taken from the total live leukocyte population. Statistical analysis here utilises the unpaired Student's T-Test.

Assessment of Ifit1 expression within the cell subtypes

Initial FACS isolation of cell populations for later qPCR analysis proved unsuccessful, as RNA levels were barely detectable across the samples. Nucleic acid concentrations ranged from 0 - 9ng/μl, with low accompanying 260:280 values. This was therefore prohibitive for taking samples forward for qPCR analysis of Ifit1 expression. One of the ways to further control this experiment would be to utilise the transfected HEK293t cells as positive controls. This lack of useable RNA led to a change in approach towards positive selection of Ifit1 containing cells by the FACS equipment.

6.4.3 FACS Analysis on Permeabilised Cells

FACS analysis was conducted on permeabilised leukocyte cells, incubated with the biotinylated anti-Ifit1 antibody. When applying the gates as originally optimised in section 6.3.7, very few cells showed a positive signal (Figure 6.16). This lack of positive signal resulted in very few cells able to pass through the various other gates used for determining leukocyte population (Figure C.1).

6.5 Discussion

6.5.1 Ifit1 Role in Hypertension

A growing body of evidence is suggesting the importance the innate immune system plays in several inflammatory processes known to affect the cardiovascular system. Innate immunity refers to the body's first line of non-specific defense mechanisms against pathogens. However, the innate immune system also plays an important part in responding to endogenous damage. Increasing research is highlighting Ifit proteins in the innate immune response, specifically the Ifit1 protein. In 2012, McDermott *et al.* used several large-scale transcriptomic datasets from mouse and mouse macrophage studies

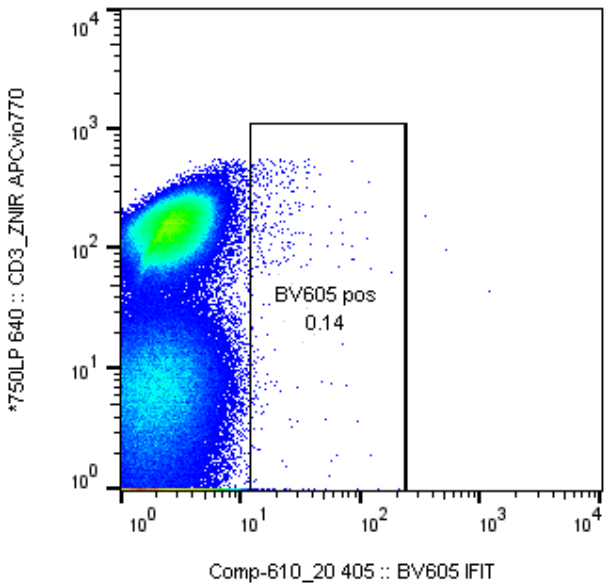


Figure 6.16: fluorescence-activated cell sorting (FACS) output of BV605 Positive cells (Stained with anti-Ifit1 antibodies), illustrating how few yielded a positive signal.

to identify a series of targets important in controlling innate immune responses.[209] By comparing these datasets, the team were able to infer networks from statistical associations between transcript expression profiles. From these networks, genes with a high-betweenness centrality were deemed important "bottlenecks" of the innate immune response. Across all the datasets analysed, they found only six conserved bottleneck genes including; Ifi47, Axud1, Ppp1r15a, Tgtp, Ifit1, and Oasl2. They continued to validate Ifit1 through qPCR analysis of expression in model of LPS-mediated TLR4 activation. Using Ifit1 siRNA expression of the predicted first-order network genes was robustly reduced, suggesting Ifit1 exerts an important regulatory influence across several downstream immune genes in this model of infection.

The method of inducing the innate immune response by way of LPS exposure is a common one. However studies on LPS and blood pressure appear to yield mixed results. Many studies into the acute effects of LPS observe a reduction in blood pressure as a result of the infection, accompanied by a steady increase in heart rate to account for the lower blood pressure.[210] Another study into the time course effects of hemodynamic responses to LPS noted a rapid reduction of blood pressure after a low dose of LPS

(0.06mg/kg) was injected. Interestingly, the same team saw a significant increase in blood pressure at doses between 20-40mg/kg.[211] Chronic studies on LPS are less revealing, as animals rapidly desensitise to the exposure of LPS and as a result blood pressure remains unchanged for the duration of the study.[212]

It would certainly be of interest to manipulate this system in future, with LPS exposure at a higher level to see the effect it has on blood pressure and whether it is accompanied with a concomitant rise in Ifit1 expression occurs within the blood.

6.5.2 Ifit1 in the paraventricular nucleus (PVN)

Here a significant upregulation of Ifit1 expression was observed within the PVN. The hypothalamic PVN has been heavily implicated in the relationship between innate immunity and hypertension. Studies have shown increased levels of pro-inflammatory cytokines (PICs) to play an important factor in the development of hypertension. Shi *et al.* (2010) show the involvement of microglial cytokines in neurogenic hypertension by administering either an anti-inflammatory antibiotic, or an adeno-associated virus for the overexpression of the anti-inflammatory IL-10 alongside angiotensin II in the PVN. The team measured a series of physiological metrics associated with a hypertensive disease state including; mean arterial pressure, cardiac hypertrophy, and plasma norepinephrine after chronic angiotensin II administration. They found a significant reduction in all of these parameters after both of the treatments as compared with PBS or AAV-GFP controls. The team further quantify a significant increase in activated microglia, as defined by increase OX42 antibody binding, the neuroglial cells tasked with innate immune responses to pathogens. Taken together these results highlight an important role for microglia and pro-inflammatory cytokines in AngII induced hypertension within the PVN ([213]). In addition to the immunological roles of these cytokines and their receptors, a large body of evidence suggests a role for them as vital messengers in CNS regions.[214] Ifit1 specifically has been observed in the PVN regions of 6 human donors

based on a search of the Allen Brain Atlas (2010 Allen Institute for Brain Science).

For this reason, coupled with its other important regulatory roles outlined above, the PVN was a key region for investigating Ifit1 involvement in hypertension.

Indeed a case could be made for Ifit1 being a consequence of toll-like receptor 4 (TLR4) signal transduction within the PVN. One of the key aspects of innate immunity is the recognition of pathogen associated molecular pattern (PAMP)s in order to exact an immediate and coordinated attack on possible pathogens. Cells do this via pathogen-recognition receptor families (PRRs), such as that of the toll-Like Receptors (TLRs). The toll surface protein was initially discovered in Drosophila and characterised for its developmental role in determining the dorsal-ventral axis in embryos.[215] Toll homologues in humans were described in the late 1990s, and became later renamed as toll-Like Receptors (TLRs). At the time, there was no knowledge of the function they play in human biology. However, today they are known for their major role of the innate immune system, an attribute which can be exemplified by how evolutionarily conserved they are between species.[216] Most vertebrates have one ortholog for each of the TLRs family, and sequencing experiments have revealed them in genomes as diverse as; chickens (*Gallus gallus*), Japanese pufferfish (*Takifugu rubripes*), and the opossum (*Monodelphis domesticus*).[217] In total, 13 isoforms of TLRs have been identified in mammalian cells, with each recognising a specific PAMP.[218]

Upon activation TLRs set off a diverse range biomolecular cascades. There are two major downstream branches of TLR activation that are characterised by the receptor adapter proteins they recruit; Myeloid Differentiation Primary Response Protein 88 (MyD88), and TIR-domain-containing Adapter Protein Inducing Interferon (TRIF). The latter leads to IRF3 activation, inducing the transcription of a broad range of ISGs; including Ifit family proteins.[219,220] It is therefore plausible that Ifit1 upregulation could be consequence of TLR4 activation, and therefore an indirect consequence of a well documented mechanism of elevated blood pressure.

Several studies have noted the pharmacological blockade of TLR4, specifically within

the PVN, to attenuate blood pressure in a genetic model of hypertension. A study in 2015 used a bilateral inhibition of TLR4 within the PVN of the SHR and measured the blood pressure of the animals.[218] Their study firstly revealed SHR rats to have a higher level of TLR4 in their PVN compared with WKY rats. Furthermore blockade of the TLR4 using a viral inhibitory peptide (VIPER) significantly lowered the blood pressure in the SHR animals and attenuated the cardiac hypertrophy often observed within SHRs. They found TLR4 inhibition to attenuate PICs (e.g. TNF-α), induced Nitric Oxide Synthase (iNOS) and the activity of the NF-$\kappa\beta$ as well as an increase to levels of anti-inflammatory IL-10 within the PVN of hypertensive rats. Along with these observations, the team found a significant reduction to circulating plasma Noradrenaline in the SHRs, a circulating hormone responsible for the constriction of blood vessels to increase blood pressure.[221] As NF-$\kappa\beta$ was seen to be significantly reduced, the team suggest this attenuation to be attributable to a reduced inflammation, in turn leading to disruption of a detrimental positive feedback cycle involved in the progression of hypertension. Indeed recent studies have shown AngII induced hypertension to be blocked by NF-$\kappa\beta$ blockade in the PVN.[222] The team continue to postulate the method by which TLR4 blockade to attenuate blood pressure as being via ultimately inhibiting sympathetic outflow. Noradrenaline is an indirect indicator of sympathetic activity, and was significantly reduced in the SHRs treated with the TLR4 blockade. Additionally, TNF-α and NF-$\kappa\beta$ have both been implicated in cardiac remodeling in hypertension and end organ damage respectively.[223,224] Strengthening this hypothesis when taken together with the team's observations on cardiac hypertrophy.

A study in 2018 built upon these findings using the TAK-242 selective blocker of TLR4 in the PVN of SD rats fed on a low (0.3%) or high-salt (8%) diet.[225] The team also noted elevated TLR4 expression in the PVN of the hypertensive rats, however this time in an environmentally induced form of hypertension. This was alongside a corresponding increase in reactive oxygen species and PICs. The team then used a different experimental method to block TLR4 expression, and again noted a resulting

reduction in blood pressure. Excessive reactive oxygen species have a well established role in increasing sympathetic nerve excitability, ultimately leading to an elevated blood pressure.[226] These findings shed additional light on the method by which TLR4 activity translates to elevated blood pressure. While these studies did not study the resulting levels of Ifit1 expression in the PVN, they do implicate the innate immune system that has been associated to "bottleneck" upon this protein. These studies highlight the need for further analysis of ifit1's central role in blood pressure regulation in the PVN and suggest TLR4 activity in the PVN as having a potential involvement in ifit1 elevated expression.

6.5.3 FACS Analysis to Isolate Cell Types

Here focus was given to resolving the leukocyte population responsible for the elevated abundance of Ifit1 mRNA in the SHR blood. Whole blood contains several types of blood cell that can be broadly grouped into; erythrocytes, leukocytes, and thrombocytes. Of these, only leukocytes contain a nucleus and are therefore the subject of interest for studying elevated transcription of a transcript. However in 2009 a revolutionary study, that making use of microarray analysis, revealed the presence of over 1000 transcripts in a population of erythrocytes.[227] This observation has clear ramifications towards studying the transcriptome of whole blood. While it should be noted that this study has not since been repeated, focus needs to be given to anucleated cells such as the erythrocytes and their progenitors (i.e. reticulocytes) when trying to deduce the source of a transcript of interest. Probing this 2009 dataset did reveal many the Ifit family of genes were detected by the microarray analysis. It may therefore be of use here to include erythrocytes in downstream analyses as they could be the source of the robustly elevated Ifit1 transcript.

Despite the high availability of rat bloods from other's experiments, the decision was made to prioritise optimisation of the above protocol in SHR blood alone. The primary

reason for this was the well documented increase in total leukocytes in SHR blood when compared to the WKY blood. A study from 1991 found an increase in leukocytes of between 50-100% in young, mature, and old hypertensive rats compared with controls.[228] The study further showed a 300% increase in activated granulocytes, detected by a nitroblue tetrazolium reduction, in the SHR. These effects appeared consistent in rats obtained from colonies in Massachusetts, Denmark and [West] Germany. Therefore mitigating the various differences related to colony and seasonal variations in leukocyte counts. For this reason, it seemed pertinent to optimise antibody dilutions for use in the SHR, so as to prevent; lack of binding due to low concentration or signal saturation due to over binding with a high concentration.

A series of interesting observations arose from the initial FACS analysis of SHR and WKY blood cells. The relative abundances of specific leukocytes were significantly different between the two strains, notably the CD43 High/His48 Low monocytes, and the natural killer cells. These experiments were conducted with only 3 replicates, however the variance seen was sufficiently low to produce significant differences between the two cohorts. The observation of a differential abundance of leukocytes between these strains is not a new one, as many studies have noted the elevated levels of total leukocytes in the SHR as compared with the WKY.

NK Cells

Interestingly, the SHR blood saw a significant increase in NK cells as compared with the WKY blood, where none were detected. Natural killer (NK) cells are non-B, non-T lymphocytes with the ability for "natural" antigen-independent cytotoxic activity and as such play a central role in the innate immune system.[229] NK cells are responsible for responding to abnormal or stressed tissue, and can activate or suppress inflammatory responses via their interactions with T-Cells, Dendritic cells, and macrophages.[230]With regards to hypertension, a mutual activation has been demonstrated between NK cells and monocytes in an angiotensin II-induced model of hypertension.[231] This connection

between NK cells and hypertension was furthered through the observations on vascular remodeling by Taherzadeh *et al*, who used a congenic strain bred from C57BL/6 and BALB/c mice to show those sharing the same NK gene complex had a similar blood pressure response to a chronic L-NAME-induced hypertension model. This study highlights the potential role for NK cells in the sensitivity towards nitric oxide synthase (NOS) inhibited forms of hypertension.

It has been shown in mice that angiotensin II induced inflammation and vascular dysfunction are associated with an accumulation of natural killer (NK) T cells in the aortic wall which, upon activation, rapidly release IFN-γ.[231] This angiotensin II caused effect is drastically reduced in $Tbx21^{-/-}$ a transcription factor that drives IFN production. Due to this massive upregulation in NK cells, coupled with the regulation of Ifit1 by IFNs, it appears quite probable they could be the source of the upregulated Ifit1 in the SHR blood. Either directly as the source of Ifit1, or indirectly through the upregulation of IFNs.

Monocytes

Monocytes also represent another rapid responder to areas of inflammatory response, where upon activation differentiate to macrophages.[232] Unlike NK cells, a great deal of work has explored the relationship between macrophages and hypertension. Macrophages are always present within the vessel walls and within the kidney in hypertensives.[233]

Via the detection of CD43 and His48, the total monocyte population appeared higher in abundance in the SHR, however this trend did not reach statistical significance (P=0.053). Based on the validation carried out in Barnett-Vanes *et al.* (2016), these two cell surface markers are sufficient to differentiate between two monocyte subtypes; classical, and non-classical.[205] In rats, the CD43 hi and lo monocytes are considered to be analogous to the mouse Ly6C Lo (non-classical) and Hi (classical) monocytes respectively.[234,235] Classical (CD43hi/His48lo) monocytes represent the most abundant subtype, and are thought to represent newly released cells from the bone marrow.[236]

The non-classical (CD43lo/His48hi) monocytes are lower in abundance at around 40% of the total circulating monocytes, until they expand in inflammatory states.[236] This subtype is implicated as playing a role in patrolling the early immune response, tissue repair, and neovascularisation, as well as an increased production of TNFα.[237] They have been dubbed the "sentinels" and "caretakers" of vascular tissue, due to their ability to recognise and clear dying endothelial cells, maintaining vascular homeostasis.[238,239] While this non-classical subtype seem to be primarily involved in a homeostatic caretaking role they have also been seen to contribute to several autoimmune and inflammatory diseases, such as; rheumatoid arthritis, multiple sclerosis, and aetherosclerosis.[240]

It was this non-classical subtype that saw a significant increase in abundance within the SHR blood, an observation that has not previously been noted in this model of hypertension. Indeed, this specific subtype has been implicated in human hypertension in a 2018 study conducted by Loperena *et al.*, who used a FACS gating strategy to quantify monocyte populations in hypertensive and normotensives. The team observed a progressive decline in classical monocyte populations, and a concomitant increase in the percentage of non-classical monocytes as levels of hypertension increased.[236]

The resulting cell populations were sorted directly into trizol, in an attempt to stabilise the cells and contained RNA as quickly as possible. Following a well established RNA extraction protocol, as outlined previously, the RNA yield was prohibitively low and prevented downstream qPCR analysis of Ifit1 expression levels. Due to time constraints, the samples for this experiment were immediately frozen at -80°C in order to stabilise them. After conducting a literature review to troubleshoot the protocol, several groups had noted a significant drop in RNA recovery following the freezing of the sample-trizol mix with some reporting as much as a 50% reduction in RNA yield.[241]

A further concern could be the presence of RNases in the vicinity of the FACS machinery. It was not feasible to thoroughly prepare the instruments to reduce RNase activity. However, some protocols suggest adding an RNase inhibitors to the FACS

suspension in order to mitigate any RNA degradation.[242] This approach presents a simple modification to the protocol followed above, and may well yield higher quality RNA for downstream analysis.

6.5.4 FACS Analysis on Permeabilised Cells

Due to the resulting RNA being of poor quality, qPCR could not shed light on the relative abundances of Ifit1 expression within the different cell populations. The experiment did however yield a novel insight into the relative abundances of leukocytes, as outlined above. In order to overcome the poor RNA yield, a change in approach was attempted. Rather than physically isolating the live cells before extracting RNA, cells were permeabilised in order to allow antibodies to access the intracellular Ifit1. This way, Ifit1 positive cells could be detected and cross-referenced with the signal profiles of cell surface markers used to discern between leukocyte types. Furthermore, the act of permeabilisation required the fixation of the cells; thereby killing them. This fixation approach did not present as many concerns towards cell/transcript stability that were present for the live-cell sorting approach.

This approach was unsuccessful in resolving Ifit1 containing cell populations, due to the low levels of Ifit1 positive cells being detected. There could be many reasons behind this, all centering on one of two possibilities; either the experimental approach failed to detect the expressed Ifit1 protein, or Ifit1 protein is not expressed at levels comparable to the transcript.

It must be noted, that here previous analysis has been conducted on transcript levels of Ifit1 within the blood and PVN. A clear and robust increase in expression is occurring between the strains being tested, however this expression may not result in different levels of functional protein. A good review on the matter was conducted by Vogel and Marcotte (2012), who showed a correlation between expression levels and resulting proteins to be as little as 40%.[243] There exist many processes of translational

control between transcript expression and function protein, all of which present a stage at which a disparity can arise. mRNA stability, Protein stability, and the rate of mRNA transcription *versus* protein translation all factor into the correlation observed.

Focusing biomarker discovery work on transcripts is a valid approach if they stably and reliably act as a metric for physiological changes, as observed here. However, no conclusions can be drawn on Ifit1's specific role within hypertension without the use of a reliable antibody that is robustly validated. Transcriptomic analysis therefore presents a useful way to identify potential candidates for down-stream work at the level of the protein.

6.5.5 In Silico Characterisation of Ifit1

Since the robust increase in ifit1 mRNA was present in the juvenile SHR at only 4 weeks, and previous blood pressure measurements of the SHR suggest it to be hypertensive at this age, it remains unclear whether Ifit1 is causative, resultant or neither with regards to hypertension.

The high sequencing depth, coupled with the relatively large number of replicates, enabled me to look towards the NGS reads to resolve and nucleotide specific changes between the WKY and SHR Ifit1 mRNA in the blood. I was therefore able to characterise the genetic differences that exist between the WKY and SHR ifit1 gene, and furthermore whether these differences are responsive for the altered expression levels. With this in mind, and using the BAMtools suite of bioinformatic tools, I was able to merge all of the previously aligned BAM files from the juvenile WKY and SHR blood RNAseq experiment (n=6). This resulted in two BAM files containing all of the reads, aligned to the RN6 reference sequence, across all WKY samples and all SHR samples for a straightforward downstream analysis. The WKY and SHR compiled BAM files where sorted and indexed using the bamtools index command. Both the .bam aligned reads and the .bam.bai indexes could be read into the Integrative Genome Viewer (Version

2.4.4, Broad Institute California).[244,245] As these were both aligned and indexed against the same reference sequence (RN6), a simple search for "ifit1" localises the viewing panel to the locus of ifit1 in the rat genome. Once all reads are visualised against the ifit1 locus, a coverage histogram could be generated. Upon first assessment, and as expected with the predicted upregulation observed in chapter 1, there is a consistently higher number of RNAseq reads aligning to the gene from the SHR samples, than in the WKY samples. What is interesting, with regards to the WKY sample reads, is the relatively short span of ifit1 gene towards the 5' that shows a high number of aligned reads (reaching 1789 overlapping reads).

The IGV software also allows for single nucleotide polymorphism calling (SNP), by comparing each read base against the reference genome. For the above coverage histogram, an allele frequency threshold can be arbitrarily set to reveal the site having that proportion of alleles called differently to the reference genome. Any loci whose proportion of reads are above this threshold will highlight as red on the coverage histogram (Table 6.14). The default threshold setting is left at 0.2, meaning over 20% of reads on that particular locus have to differ from the reference base for the locus to highlight red. To increase stringency, and assess how prevalent any discovered SNPs are within each strain, I iteratively increased this threshold to inspect the resulting coverage profiles. At a threshold of 1.0, meaning an differential base calling event in every single read within the SHR population, 3 loci were highlighted as being different.

Locus	Percentage of reads	SNP	Amino Acid Change	Intron/Exon
Chr1:252,944,180	100%	C → T	S → L	Exon
Chr1:252,944,345	87%	C → T	S → L	Exon
Chr1:252,944,471	99%	A → G	N → S	Exon
Chr1:252,944,488	100%	A → G	T → A	Exon
Chr1:252,945,968	100%	G → T	R → I	Exon

Table 6.14: Location of single nucleotide polymorphisms in SHR Ifit1 transcripts, as compared against the RN6 reference genome

However, it must be noted that as there were no reads from the WKY samples covering the above loci, it is not possible to assess whether these SNPs were truly part

of a different genetic profile in the SHR. They may well be a part of a genetic profile shared between the strains, yet different from the RN6 reference genome the bases are being compared against. When this threshold was dropped to 0.87 (87% of reads), two more loci were highlighted. What is interesting is that one of these aligns with the only section of the ifit1 gene that saw a large number of reads within the WKY samples. At Chr1:252,944,345, an exchange of Cytosine → Thymine was observed in the SHR across those 87% of reads covering that locus. For the same site for the WKY reads, no such SNP has taken place. This is therefore the only site where evidence exists to suggest a SNP existing between the WKY and SHR at this site.

At this point it should be noted that the sequence-specific error profile of RNAseq data is in no way 100% accurate. This is a major consideration when attempting *de novo* assembly or indeed the SNP calling attempted here. The potential issues with RNAseq inaccuracy are discussed earlier in greater length, however SNP calling is most reliant on preventing substitution-type miscalls of bases. A comparison study in 2011 looked into the various different sequencing platforms and estimated miscall rates to be $\geq$0.1% of bases. This translates to a conservative 1 miscalled base in every 1000 bases sequenced.[245] While this number can be greater depending on platform used to sequence, our laboratory uses a Phred cutoff score of >Q30, translating to 0.1% miscalls per base. This threshold ensures the samples included in our analyses are at what is considered to be the current benchmark of accuracy given the library preparation chemistry at the time of sequencing. Illumina purports sequencing phred scores can often exceed this score, however it is seldom reported in the literature. Another consideration with base miscalls and phred score, is that this is calculated as an average of the entire read length. This becomes a problem because the miscall error rate tends to increase along the read length, as outlined in section 1.5.2 (Figure 1.6).[246] For this reason, a read with a Phred Q30 score could on the surface appear sufficient to pass the stringent exclusion criteria for a given study. However, if this is an average of a read of Q40 between bases 1-100 and Q10 for the bases 101-150, the latter section of the read has

a significantly large error rate of 1 miscall per 10 bases. This should therefore be a consideration when evaluating RNAseq data, and a level of stringency should exist that is sufficient to mitigate these error rates. Since these SNPs appear in >80% of reads, coupled with a high read depth, for each loci it seems highly unlikely that these could be miscalls. Probing the ENSEMBL database for Ifit1 reveals 25 SNPs have been identified within the region. All 5 of the SNPs detected here have also been listed as part of this database however none of which were linked with any publications discussing their potential importance. Further assessment of these potentially different SNPs could be achieved by PCR amplification of the ifit1 region, before Sanger sequencing to reveal any differences in the ifit1 gene between the WKY and SHR strains. Due to the implication of many disease states with SNPs, looking towards these polymorphisms could prove an important direction to take for predicting disease susceptibility.

Chapter 7

Reflections & Future Directions

7.0.1 Furthering the validation of Ifit1

This body of work has laid a strong foundation in furthering the discovery of novel biomarkers of hypertension. Using an optimised approach towards RNAseq analysis, a candidate list was produced of transcripts that were differentially expressed in the blood between the WKY and SHR strains. From this list several candidates were taken forward and validated using independent experimental approaches and additional models of hypertension. However, one of the key questions remaining unanswered is the specific aetiology of hypertension that Ifit1 expression is indicative of. It is clear that Ifit1 is abundantly expressed in a genetic model of hypertension, unlike any of its other protein family members. However when attempting to environmentally induce hypertension Ifit1 saw a counterintuitive decline in expression. It follows that Ifit1 may simply be a metric for genetic predisposition, and not one of blood pressure in its entirety. Here, only two genetic models of hypertension were assessed; the SHR and the BHR. Both of which are so closely related that it is not possible to discern whether Ifit1 upregulation is a relfection of the SHR lineage or indeed a true biomarker of genetic hypertension. Several genetically distinct models of hypertension exist, as shown in figure 5.4. By examining the expression of Ifit1 in these strains, it should be possible

to conclude whether or not Ifit1 is indeed a metric of hypertension or simply the SHR strain itself. In order to further reconcile this uncertainty, efforts should be made to obtain the Dahl-Salt sensitive and salt resistant strains for analysis. This model of hypertension would allow for an assessment of the balance between environmentally induced and genetically predisposed forms of elevated blood pressure. Another area to explore would be the role of ifit1 in response to certain immune challenges. Due to its well characterised involvement in the innate immune response, its elevated expression in a human population may be indicative of certain immune challenges. It is therefore likely that expression levels may need to be assessed in conjunction with other non-immune responsive genes to ensure its expression is not resulting on infection. It would be possible to mitigate this concern by looking solely are the genomic level for features associated with hypertension, such as SNPs. There exists a wealth of literature studying the correlations between polymorphisms and disease states. However genomic studies will only show genetic predisposition towards hypertension, and not the contribution of the many lifestyle factors established to lead to elevated blood pressure. It was therefore the aim of this current body of work to attempt to look towards the transcriptome as a way of reconciling this genome-environment interface.

Furthermore, due to the high-throughput nature of this study there exist many other candidate biomarkers that may well be successfully taken through the various validation strategies outlined here. Herein lies the novel benefit of RNAseq technologies; to guide biological discovery in a manner free from bias. As more studies are produced, a greater level of data mining can occur to elucidate important differences in gene expression, allowing for an increasingly guided selection of candidate biomarkers to be made.

7.0.2 The Aetiology of Hypertension

This study set out with the aim of identifying novel prognostic biomarkers of hypertension. Although a biomarker only needs to present an objective measure of a disease

state, it would be of interest to further characterise any mechanistic involvement in hypertension, if at all. By exploring the biology of these biomarkers in the context of blood pressure regulation it may enable us to shed light on how hypertension arises.

While results were inconclusive, preliminary work has been carried out to ascertain Ifit1's involvement in hypertension by characterising the leukocyte population responsible for its expression. Future research should therefore build upon this to enable a source to be identified such that a more guided literature review of leukocyte involvement in hypertension be conducted.

7.0.3 Translating the findings to a human cohort

From these findings it is now of great interest to expand validation studies to assess how suitable these biomarkers are for a metric of human hypertension. With this in mind, the quality in organ donation (QUOD) initiative presented a novel opportunity to test biomarkers in human tissues from patients with comprehensive medical histories noted.

The initiative was formed as part of the Nuffield Department of Surgical Sciences with the aim of identifying pathways of injury and repair in donor organs, in the hopes of mitigating any arising complications to organ recipients. It is well known that patients of donated organs from these with high blood pressure exhibit a lower survival rate.[247] This present study aims to see if novel changes taking place in the blood can help to identify those that have a higher blood pressure. This would enable improved screening of postmortem donors where it is no longer possible to measure blood pressure for the biomarkers associated with chronically elevated blood pressure as identified here. Furthermore, this study is also looking towards prognostic changes that occur before the onset of high blood pressure. This would allow for treatment plans to be put in place prior to blood pressure rising and putting strain on internal organs.

Due to the operating of the initiative, blood tissue was not available from donors. However, as the spleen is heavily vascularised it will be used as a proxy measurement

of blood transcripts. At the time of writing, ethical approval was granted for usage of these tissues for biomarker validation. From an initial list of 364 potential male donors, 40 individual samples have been selected based on the clinical data provided in order to conduct a pilot study. This was carried out by narrowing down the total potential donor list to those with no history of smoking, or cardiovascular disease and binned into those with documented hypertension and those with no history of hypertension. From this narrowed list of potential candidates 20 patients were selected from either the hypertensive or normotensive cohorts based on how comprehensive their medical history was.

Tissues have since been received from the QUOD biobank and are currently stored at -80°C ready for RNA extraction and quantification. Linear regression modelling will then enable us to assess whether any correlations between expression and blood pressure exist. This will allow the direct assessment of how suitable our biomarkers are in a human population, and to further assess how well transcripts obtained *post mortem* correlate with patient blood pressure *pre-mortem*. It therefore presents an interesting avenue for furthering research into biomarkers of hypertension.

7.0.4 Concluding remarks

Given the significant health and financial impact of essential hypertension, coupled with an ever increasing population, work towards a better diagnostic approach is of huge importance. Only with continued study towards identifying the aetiology of hypertension will it result in improved diagnostic assessment, ultimately leading to a therapeutic design centred on prophylaxis rather than symptomatic management. With this in mind, High-throughput 'omics' approaches hold a great deal of promise in providing a thorough characterisation of disease states. In diagnosing patients earlier in the disease's progression, sufficient preventative steps can be put in place to minimise the burden placed on healthcare, and ultimately an individual's health.

Appendix A

Chapter 1

		Percentage of Aligned Reads (%)			
		Reads Aligned	Multiple Alignments	Discordant Rate	Concordant Rate
	WKY1	66.7	0.4	62.8	23.2
	WKY2	67.5	0.4	60.7	25.1
	WKY3	63.3	0.7	59.5	24.5
	SHR1	68.2	0.9	44.5	35.1
	SHR2	68.1	1.2	41.3	37.3
	SHR3	61.7	1.3	45.5	31.9
Mean	**All**	**65.92**	**0.85**	**52.38**	**29.52**
	WKY	**65.83**	**0.57**	**61.00**	**24.27**
	SHR	**66.00**	**1.13**	**43.77**	**34.77**

Table A.1: Post-Trimming Alignment Rates for n=3 RNAseq Dataset

				DESeq		EdgeR		DESeq2	
	GeneName	SHR.Avg	WKY.Avg	FC	P-Value	FC	P-Value	Log2FC	P-Value
ENSRNOG00000000033	Tmcc2	64736.4	7594.7	0.117317	4.79E-62	0.111746	2.74E-62	-3.06493	2.29E-151
ENSRNOG00000038999	RT1-A1	1238.02	133.756	0.10804	3.49E-42	0.103461	8.16E-43	-3.13417	5.84E-70
ENSRNOG00000021166	Ecm1	374.024	31.1632	0.0833187	3.56E-27	0.0790522	6.64E-38	-3.42122	2.58E-35
ENSRNOG00000001242	Gstt3	132.702	0	0	1.60E-22	0.00068279	2.78E-28	-4.99751	2.74E-18
ENSRNOG00000013288	Xpo7	10214.5	2527.73	0.247466	4.40E-26	0.235607	7.51E-27	-1.99563	1.98E-54
ENSRNOG00000019050	Ifit1	2324.18	296.53	0.127585	1.30E-45	0.121078	1.33E-25	-2.90071	1.70E-47
ENSRNOG00000036603	Ifit1bl	1008.68	9801.03	9.71666	1.78E-53	9.23328	3.27E-24	3.18901	7.61E-51
ENSRNOG00000025108	Ankrd35	91.2945	0.603588	0.00661143	2.16E-14	0.008237	4.42E-23	-4.35357	3.09E-14
ENSRNOG00000020541	Nprl3	361.833	67.911	0.187686	6.90E-15	0.177128	8.23E-20	-2.34012	2.56E-21
ENSRNOG00000014399	Aamp	245.838	38.3486	0.155991	7.66E-13	0.150882	2.70E-19	-2.52306	4.50E-18
ENSRNOG00000039776	Exoc3	54.7416	282.878	5.16753	4.33E-12	4.93132	3.97E-18	2.29228	2.42E-21
ENSRNOG00000000451	RT1-Ba	8.33442	83.0238	9.96155	3.68E-08	9.32351	6.97E-18	3.00482	6.68E-15
ENSRNOG00000008656	Snca	502.49	1944.64	3.87	1.02E-16	3.68441	1.04E-17	1.92232	7.95E-33
ENSRNOG00000021713	Kif18b	146.18	21.6502	0.148106	3.02E-09	0.138806	1.51E-17	-2.56471	6.20E-13
ENSRNOG00000011296	Cenpn	23.006	230.193	10.0058	8.20E-19	9.48718	1.99E-17	3.11501	2.91E-22
ENSRNOG00000014289	Arpc2	1261.43	365.142	0.289468	2.18E-15	0.27522	3.13E-17	-1.76915	3.14E-29
ENSRNOG00000046607	Cited4	46.3356	0.603588	0.0130264	1.22E-06	0.0162557	1.27E-16	-3.52403	NA
ENSRNOG00000011422	Atmin	11.1536	78.5278	7.0406	4.35E-06	6.65679	1.37E-16	2.58325	3.65E-12
ENSRNOG00000008165	Tpx2	577.196	181.051	0.313672	8.33E-10	0.29802	1.46E-16	-1.64252	4.23E-19
ENSRNOG00000022911	Hjurp	77.2328	6.08149	0.0787423	3.53E-07	0.0785578	6.24E-16	-3.00514	1.21E-09
ENSRNOG00000012067	Fam111a	201.272	49.5143	0.246007	2.97E-07	0.233334	1.15E-15	-1.94002	2.91E-12
ENSRNOG00000020551	Bles03	178.254	603.874	3.38772	8.11E-10	3.21682	1.43E-15	1.72765	1.80E-23
ENSRNOG00000012722	Ppdpf	54.7026	338.566	6.18922	1.37E-14	5.8424	1.64E-15	2.51355	1.49E-22
ENSRNOG00000000150	Dnajc1	95.068	359.204	3.77839	3.01E-09	3.58478	9.06E-14	1.86595	4.19E-18
ENSRNOG00000057706	Kdm5c	73.3491	250.964	3.4215	2.01E-06	3.24526	1.38E-13	1.72191	8.44E-15
ENSRNOG00000045626	Pkd1l2	70.5025	9.02805	0.128053	4.19E-05	0.124287	2.03E-13	-2.52506	1.05E-07
ENSRNOG00000025295	Mavs	47.6126	228.053	4.78977	5.91E-09	4.51235	3.51E-13	2.15566	1.02E-15
ENSRNOG00000007546	Asns	405.293	1122.09	2.76859	2.11E-08	2.63237	4.16E-13	1.44921	5.44E-23
ENSRNOG00000018009	Rab8b	77.2056	333.169	4.31535	1.38E-09	4.07913	1.04E-12	2.03056	1.25E-17
ENSRNOG00000017259	Tacc3	246.029	55.9484	0.227406	3.52E-09	0.216611	1.14E-12	-2.0412	1.06E-12
ENSRNOG00000017932	St3gal2	450.276	1163.39	2.58373	1.32E-07	2.45864	1.46E-12	1.35351	9.56E-22
ENSRNOG00000032708	RT1-Bb	21.5336	101.49	4.7131	6.93E-05	4.42516	2.30E-12	2.08258	5.35E-10
ENSRNOG00000023303	Dpep2	34.0444	148.882	4.37318	1.48E-06	4.1869	2.52E-12	2.02471	2.19E-11
ENSRNOG00000018783	Bcas2	2016.82	5176.11	2.56647	3.32E-09	2.4441	4.17E-12	1.34607	1.50E-24
ENSRNOG00000013255	Dnajc8	150.052	500.422	3.33499	7.31E-09	3.16493	8.94E-12	1.69539	4.67E-17
ENSRNOG00000004382	Arel1	161.467	463.602	2.8712	8.32E-07	2.72748	1.20E-11	1.49159	5.80E-16
ENSRNOG00000014605	Lig4	32.4681	142.982	4.40375	7.29E-06	4.15123	2.08E-11	2.01533	4.41E-11
ENSRNOG00000021802	Isg15	58.4428	264.369	4.52356	2.36E-09	4.27194	2.25E-11	2.07569	5.58E-14
ENSRNOG00000033765	Eif1	78.9022	306.283	3.8818	2.43E-08	3.69042	3.01E-11	1.88323	3.45E-14
ENSRNOG00000003385	Glrx2	14.6512	86.849	5.92777	6.47E-06	5.66945	3.76E-11	2.33282	7.25E-09
ENSRNOG00000026542	Suco	39.7531	147.678	3.71488	7.13E-05	3.51275	4.10E-11	1.80151	1.96E-10
ENSRNOG00000051682	Zcchc9	37.2993	2.10571	0.0564544	0.00052511	0.0557766	5.67E-11	-2.81946	NA
ENSRNOG00000017720	Mxd1	486.964	1172.09	2.40693	9.81E-07	2.29358	1.33E-10	1.25292	1.17E-17
ENSRNOG00000048277	Srrt	88.2075	21.7826	0.246947	0.00130734	0.23529	1.58E-10	-1.84303	2.81E-06

Table A.2: Original RNAseq Output of WKY vs SHR 4 weeks blood comparison

			DESeq			EdgeR		DESeq2	
	GeneName	SHR.Avg	WKY.Avg	FC	P-Value	FC	P-Value	Log2FC	P-Value
ENSRNOG00000047314	Tk1	680.296	1753.53	2.5776	5.51E-08	2.45202	1.83E-10	1.34738	3.25E-19
ENSRNOG00000017079	Pgm2l1	18.1505	77.3784	4.26315	0.00122055	4.06463	1.99E-10	1.93619	1.46E-07
ENSRNOG00000019934	Paip2	352.341	983.312	2.7908	3.32E-08	2.65019	2.40E-10	1.45441	1.17E-17
ENSRNOG00000027089	Ell2	1132.92	2727.74	2.40771	2.64E-07	2.2926	2.68E-10	1.25408	1.21E-21
ENSRNOG00000023202	Usp15	1502.16	3395.25	2.26025	8.09E-07	2.15122	2.82E-10	1.16633	5.07E-21
ENSRNOG00000009495	Src	96.6782	20.9417	0.216613	0.000194273	0.208307	3.01E-10	-1.99584	7.19E-07
ENSRNOG00000011353	Gfi1b	425.665	150.335	0.353178	9.81E-07	0.337003	3.59E-10	-1.4588	1.66E-11
ENSRNOG00000058909	Rps9	413.527	156.633	0.378773	3.32E-06	0.358749	4.41E-10	-1.37695	5.70E-11
ENSRNOG00000024605	Fam220a	368.321	873.635	2.37194	4.87E-06	2.25708	5.85E-10	1.23059	6.87E-16
ENSRNOG00000006420	Rbm38	5092.56	15599.2	3.06313	1.24E-12	2.92182	7.63E-10	1.58815	3.64E-21
ENSRNOG00000020875	Celf3	46.736	7.15797	0.153158	0.00430255	0.144752	7.90E-10	-2.22752	3.48E-05
ENSRNOG00000006833	Rb1cc1	283.392	684.15	2.41415	1.53E-05	2.29539	8.02E-10	1.25113	4.23E-15
ENSRNOG00000004742	Chp1	250.012	643.467	2.57374	1.84E-06	2.44667	9.49E-10	1.34152	6.07E-15
ENSRNOG00000010765	Vcl	249.335	89.6753	0.359658	0.000113956	0.344734	1.01E-09	-1.41588	3.15E-08
ENSRNOG00000002829	Ppbp	140.698	28.6928	0.203933	8.91E-07	0.192116	1.27E-09	-2.1236	1.03E-08
ENSRNOG00000040281	Zc3h10	24.616	102.562	4.16648	9.31E-05	4.00846	1.67E-09	1.92366	7.37E-08
ENSRNOG00000018517	Trim21	189.072	623.223	3.29622	4.18E-10	3.12966	2.17E-09	1.67528	3.26E-14
ENSRNOG00000033979	Gm5471	45.6199	149.033	3.26684	0.000347706	3.09175	2.54E-09	1.62615	1.50E-08
ENSRNOG00000011222	Dynll1	199.643	518.885	2.59906	8.39E-06	2.46735	3.18E-09	1.34902	4.35E-13
ENSRNOG00000053260	Lilrb3l	71.9022	13.4021	0.186393	0.000309199	0.170779	3.26E-09	-2.16383	2.74E-06
ENSRNOG00000000989	Bud31	100.691	313.577	3.11424	1.06E-06	2.96261	3.62E-09	1.59151	1.59E-11
ENSRNOG00000002866	Rassf6	1.76429	27.0385	15.3255	0.000875109	13.4285	3.71E-09	2.82725	NA
ENSRNOG00000010353	Dnajb6	159.153	407.993	2.56353	5.93E-05	2.43595	4.14E-09	1.32619	6.58E-12
ENSRNOG00000010081	Tmem144	13.3325	62.0045	4.65065	0.00205841	4.4054	4.76E-09	2.00626	1.37E-06
ENSRNOG00000003567	Ppox	131.475	37.6599	0.286441	0.000264949	0.271715	5.49E-09	-1.69046	1.24E-06
ENSRNOG00000059852	Rn5-8s	7.8841	82.6219	10.4796	4.23E-08	9.83915	1.13E-08	2.85026	9.83E-09
ENSRNOG00000018824	Slc7a5	371.161	982.49	2.64707	1.91E-07	2.52279	1.14E-08	1.37924	9.32E-16
ENSRNOG00000022839	Ifit3	57.8983	189.279	3.26917	1.52E-05	3.12281	1.29E-08	1.6454	2.72E-09
ENSRNOG00000022800	Sp140	295.808	99.4123	0.33607	1.22E-05	0.322919	1.64E-08	-1.49773	4.86E-09
ENSRNOG00000017123	B2m	6691.13	14394.4	2.15126	3.32E-06	2.04658	1.91E-08	1.09529	1.16E-18
ENSRNOG00000018487	Slc3a2	752.569	1706.12	2.26706	3.84E-06	2.15979	2.20E-08	1.1654	1.67E-14
ENSRNOG00000029939	Gypc	176.927	61.4006	0.347038	0.000264949	0.325852	2.43E-08	-1.47932	3.10E-07
ENSRNOG00000015762	Abtb1	987.893	2384.6	2.41382	3.21E-07	2.29956	3.65E-08	1.25389	6.20E-17
ENSRNOG00000005195	Cst3	107.225	336.563	3.13886	1.03E-06	2.97693	4.31E-08	1.59094	2.64E-10
ENSRNOG00000010993	Dpm1	33.1319	116.175	3.50645	0.00162362	3.29666	4.52E-08	1.69202	1.47E-07
ENSRNOG00000007842	Aup1	348.838	819.749	2.34994	1.61E-05	2.23707	5.30E-08	1.21203	1.36E-13
ENSRNOG00000017852	Nars	227.262	85.3739	0.375663	0.00032204	0.358119	6.54E-08	-1.36143	3.87E-07
ENSRNOG00000010331	Ctsb	12798.4	6728.4	0.525722	3.86E-05	0.500127	6.60E-08	-0.918648	4.46E-12
ENSRNOG00000020941	Psenen	700.376	1611.35	2.3007	5.09E-06	2.1894	6.99E-08	1.18418	2.58E-14
ENSRNOG00000011279	Cmc2	34.4504	131.482	3.81656	8.21E-05	3.60981	8.22E-08	1.80734	1.01E-07
ENSRNOG00000015948	Slc1a5	466.724	209.684	0.449268	0.000194815	0.426742	9.11E-08	-1.13098	3.52E-08
ENSRNOG00000008825	Ssrp1	276.33	103.355	0.374029	6.38E-05	0.355454	1.11E-07	-1.37752	3.61E-08
ENSRNOG00000033921	Als2cl	256.055	96.8486	0.378234	0.000269122	0.362115	1.25E-07	-1.33966	6.89E-07
ENSRNOG00000007806	Arf5	2138.49	4646.46	2.17278	4.25E-06	2.07202	1.27E-07	1.10648	7.88E-15

Table A.3: Original RNAseq Output of WKY vs SHR 4 weeks blood comparison Cont.

Appendix B

Chapter 3

		DESeq2			
	GeneName	Log_2 Fold Change	P-Value (Adjusted)	Mean Counts - WKY	Mean Counts - SHR
ENSRNOG00000000033	Tmcc2	3.50	2.93E-205	47532.31	537116.59
ENSRNOG00000021166	Ecm1	3.68	1.56E-181	832.70	10644.15
ENSRNOG00000048951	LOC100364500	-5.33	5.05E-165	819.73	20.36
ENSRNOG00000019050	Ifit1	7.37	7.17E-148	73.56	12195.78
ENSRNOG00000061943	Rn60_7_1214.2	-7.72	8.27E-133	938.38	4.65
ENSRNOG00000000723	RT1-CE5	-2.69	2.93E-130	11059.55	1717.98
ENSRNOG00000031090	RT1-CE7	-2.55	3.88E-109	7438.51	1273.01
ENSRNOG00000014289	Arpc2	2.10	1.26E-96	4712.04	20257.16
ENSRNOG00000012722	Ppdpf	-2.72	1.59E-94	7235.73	1098.27
ENSRNOG00000001242	Gstt3	5.11	3.28E-94	25.46	883.96
ENSRNOG00000013288	Xpo7	2.39	2.85E-81	4908.30	25666.85
ENSRNOG00000020541	Nprl3	2.06	7.89E-81	772.35	3214.74
ENSRNOG00000043176	RGD1565355	-10.56	1.12E-72	1556.47	1.12
ENSRNOG00000045626	Pkd1l2	2.92	1.12E-72	84.98	645.62
ENSRNOG00000021713	Kif18b	2.54	2.34E-66	227.62	1319.50
ENSRNOG00000014399	Aamp	2.30	6.32E-66	466.16	2289.11
ENSRNOG00000011296	Cenpn	-2.73	3.90E-64	838.26	126.33
ENSRNOG00000060864	AABR07059372.1	7.87	1.18E-56	1.94	456.34
ENSRNOG00000022911	Hjurp	3.48	5.82E-56	52.64	586.58
ENSRNOG00000046607	Cited4	4.73	2.58E-55	20.42	543.01
ENSRNOG00000045989	Hba-a3	-8.49	5.35E-55	40469.89	112.92
ENSRNOG00000020901	S1pr5	-2.56	8.87E-55	829.47	141.06
ENSRNOG00000050210	RT1-CE10	-2.12	7.42E-52	3604.14	826.99
ENSRNOG00000028415	Cdc20	-1.86	6.18E-50	2363.42	649.93
ENSRNOG00000006420	Rbm38	-1.76	8.51E-50	245847.65	72562.22
ENSRNOG00000005668	Ndufa8	1.85	2.44E-45	349.23	1261.01
ENSRNOG00000039776	Exoc3	-2.09	3.90E-45	1194.31	280.05
ENSRNOG00000019440	Kcnn4	1.67	4.86E-44	716.54	2273.52
ENSRNOG00000054978	Hist1h1c	1.72	7.09E-43	421.23	1385.38
ENSRNOG00000017259	Tacc3	1.81	3.37E-41	560.62	1966.85
ENSRNOG00000008656	Snca	-1.62	1.40E-40	19988.25	6511.90
ENSRNOG00000005195	Cst3	-2.12	4.01E-40	9454.21	2176.66
ENSRNOG00000000989	Bud31	-1.79	1.19E-39	8264.83	2390.62

Table B.1: DESeq2 generated output for n=6 differential gene expression analysis of WKY juvenile blood vs SHR juvenile blood at 4 weeks of age. Genes have been ranked by P-Value (With Benjamini-Hochberg multiple test correction), smallest to largest. Shown here are gene names, with associated ensembl gene ID accessions, alongside the Log_2 fold change between WKY and SHR. Also shown are the mean DESeq2 normalised counts for WKY and SHR.

		DESeq2			
	GeneName	Log_2 Fold Change	P-Value (Adjusted)	Mean Counts - WKY	Mean Counts - SHR
ENSRNOG00000025295	Mavs	-2.06	1.45E-39	2071.13	497.83
ENSRNOG00000008165	Tpx2	1.94	1.75E-39	543.49	2092.49
ENSRNOG00000020551	Bles03	-1.72	3.73E-39	7066.87	2146.52
ENSRNOG00000010993	Dpm1	-1.72	7.81E-39	1439.93	435.95
ENSRNOG00000061998	RT1-CE9-ps1	-3.02	9.01E-39	314.72	39.05
ENSRNOG00000042905	RT1-T24-4	-1.83	1.53E-38	6928.03	1948.03
ENSRNOG00000060821	Rnf114	-2.70	2.43E-38	2483.33	382.84
ENSRNOG00000042787	Fam229a	-2.48	3.39E-38	433.07	77.69
ENSRNOG00000029939	Gypc	1.82	1.55E-37	923.94	3263.69
ENSRNOG00000004206	Glrx5	-1.34	4.06E-37	51025.59	20203.56
ENSRNOG00000015948	Slc1a5	1.60	5.16E-36	1332.40	4026.95
ENSRNOG00000017932	St3gal2	-1.69	1.48E-35	7184.61	2226.44
ENSRNOG00000012067	Fam111a	1.78	3.37E-35	333.65	1146.89
ENSRNOG00000030712	RT1-A2	-1.77	1.06E-34	2172.61	639.38
ENSRNOG00000009282	Nub1	-1.62	1.70E-34	1568.26	510.65
ENSRNOG00000018517	Trim21	-1.96	3.70E-34	1809.59	464.38
ENSRNOG00000013255	Dnajc8	-1.57	6.34E-33	4736.91	1594.71
ENSRNOG00000013541	Sh2d4a	2.69	2.24E-32	61.57	398.22
ENSRNOG00000003385	Glrx2	-1.74	2.00E-31	1028.73	308.41
ENSRNOG00000033921	Als2cl	1.49	4.99E-31	657.55	1841.70
ENSRNOG00000022800	Sp140	1.52	1.58E-30	961.58	2751.51
ENSRNOG00000011222	Dynll1	-1.21	1.88E-30	8817.81	3807.49
ENSRNOG00000023303	Dpep2	-1.59	2.09E-30	2021.66	673.13
ENSRNOG00000002866	Rassf6	-4.00	2.15E-30	244.72	15.56
ENSRNOG00000004628	Dazap2	-1.20	3.45E-30	13687.26	5943.23
ENSRNOG00000016398	Clstn1	-3.19	5.32E-30	282.19	30.71
ENSRNOG00000017852	Nars	1.65	6.98E-30	391.86	1230.81
ENSRNOG00000011279	Cmc2	-1.47	4.34E-29	2615.06	944.50
ENSRNOG00000003567	Ppox	1.61	2.23E-28	613.22	1870.92
ENSRNOG00000010574	Ptpn1	-1.48	2.65E-28	1659.68	594.03
ENSRNOG00000040281	Zc3h10	-2.24	1.28E-27	430.66	91.41
ENSRNOG00000048924	Islr	-4.91	3.57E-27	152.29	4.99
ENSRNOG00000057706	Kdm5c	-1.38	3.74E-27	1039.26	398.43

Table B.2: DESeq2 generated output for n=6 differential gene expression analysis of WKY juvenile blood vs SHR juvenile blood at 4 weeks of age. Genes have been ranked by P-Value (With Benjamini-Hochberg multiple test correction), smallest to largest. Shown here are gene names, with associated ensembl gene ID accessions, alongside the Log_2 fold change between WKY and SHR. Also shown are the mean DESeq2 normalised counts for WKY and SHR. Cont.

					Microarray			
ID	SHR Avg (Log_2)	WKY Avg (Log_2)	SHR SD	WKY SD	Fold Change	P-Val	FDR P-Val	Gene Symbol
TC0100003518.rn.1	19.93	5.86	0	0.58	17179.2	2.31E-14	1.59E-09	Ifit1
TC0300001269.rn.1	19.92	13.24	0.01	0.33	102.5	1.49E-13	5.14E-09	plawmer
TC0200004216.rn.1	14.56	7.9	0.24	0.3	101.49	1.48E-12	3.39E-08	Ecm1
TC0X00001160.rn.1	19.2	12.7	0.26	0.45	90.35	2.84E-11	4.60E-07	gubey
TC1000000277.rn.1	10.25	7.12	0.19	0.15	8.72	3.34E-11	4.60E-07	Nprl3
TC0400000282.rn.1	19.13	13.14	0.14	0.58	63.7	1.04E-10	1.19E-06	sypar
TC0100001130.rn.1	19.93	15.39	0	0.48	23.27	2.79E-10	2.74E-06	klorfloy
TC1000001358.rn.1	10.64	16.73	0.56	0.19	-68.03	4.00E-10	3.33E-06	spumee
TC2000000588.rn.1	18.89	13.79	0.37	0.36	34.35	4.35E-10	3.33E-06	sleedy
TC0400001341.rn.1	19.6	14.12	0.17	0.6	44.84	5.96E-10	4.11E-06	deyzee
TC0900000049.rn.1	19.93	14.85	0	0.59	33.76	6.59E-10	4.12E-06	glanoy
TC1700001126.rn.1	19.93	13.54	0	0.72	83.86	8.64E-10	4.37E-06	gloree; spyklu; sweygo; warry
TC0200002591.rn.1	6.77	4.84	0.17	0.13	3.8	9.67E-10	4.37E-06	Zcchc9
TC1500000783.rn.1	14.41	8.8	0.42	0.45	48.69	9.71E-10	4.37E-06	NCBI_Gene:100861616
TC0100000567.rn.1	13.43	9.98	0.34	0.19	10.98	1.01E-09	4.37E-06	kleyfly; muga; nazer
TC0400002251.rn.1	19.91	16.11	0.09	0.44	14.02	1.01E-09	4.37E-06	klowor
TC1500001652.rn.1	11.86	8.67	0.26	0.29	9.15	1.28E-09	4.93E-06	toyja
TC0800001906.rn.1	19.93	13.76	0	0.75	72.14	1.36E-09	4.93E-06	mypoy
TC1500001518.rn.1	19.93	13.76	0	0.75	72.14	1.36E-09	4.93E-06	slorju
TC0X00001839.rn.1	19.93	15.18	0	0.61	26.91	1.55E-09	5.26E-06	verbaw
TC0900001736.rn.1	19.93	16.9	0	0.36	8.19	1.60E-09	5.26E-06	sopy
TC1800001736.rn.1	19.93	15.22	0	0.61	26.28	2.17E-09	6.80E-06	blerfaw; bleyfaw; klyter
TC0300001741.rn.1	7.18	9.88	0.27	0.17	-6.52	2.31E-09	6.92E-06	Mavs
TC0100001038.rn.1	8.09	6.65	0.1	0.15	2.72	2.56E-09	7.06E-06	Kcnn4
TC0100003809.rn.1	15.09	11.35	0.16	0.44	13.34	2.57E-09	7.06E-06	smerklaw
TC0500001234.rn.1	8.62	5.87	0.28	0.18	6.7	2.79E-09	7.39E-06	Stil
TC0900002112.rn.1	16.67	11.31	0.27	0.65	41	3.30E-09	8.41E-06	voylar
TC0200002756.rn.1	19.93	14.93	0	0.68	32.08	3.57E-09	8.47E-06	darshor
TC1500001903.rn.1	19.93	14.93	0	0.68	32.08	3.57E-09	8.47E-06	shujar
TC0700002041.rn.1	15.27	9	0.38	0.76	76.76	4.98E-09	1.14E-05	rasy

Table B.3: Microarray generated output for n=6 differential gene expression analysis of WKY juvenile blood vs SHR juvenile blood at 4 weeks of age. Here, differential expression analysis was conducted on Robust Multiarray Averaging (RMA) normalised microarray intensities using Thermofisher's Transcriptome Analysis Console (TAC, ThermoFisher Scientific). Displayed are the Affymetrics $RTA_1.0$. probe accession IDs, the average Log_2 transformed intensities for SHR and WKY samples, the corresponding standard deviations (SD), the fold change from WKY as the control group, the associated P-Value, the multiple test corrected P-Value (FDR), and the Gene name for each transcript. Transcripts have been ranked by False Discovery Rate normalised P-Value (Smallest to largest) to correct for multiple testing, before only including protein-coding genes.

					Microarray			
ID	SHR Avg (Log_2)	WKY Avg (Log_2)	SHR SD	WKY SD	Fold Change	P-Val	FDR P-Val	Gene Symbol
TC0900001920.rn.1	18.72	12.72	0.49	0.62	64.4	5.23E-09	1.16E-05	smorpu
TC0300005103.rn.1	9.41	10.75	0.14	0.12	-2.53	5.50E-09	1.18E-05	Dpm1
TC2000001805.rn.1	6.69	5.27	0.14	0.11	2.67	6.18E-09	1.29E-05	Gstt3
TC1000002799.rn.1	16.69	12.53	0.29	0.49	17.93	6.62E-09	1.34E-05	tana
TC0200001684.rn.1	6.94	5.38	0.2	0.07	2.95	7.04E-09	1.39E-05	Celf3
TC0300000315.rn.1	7.9	5.57	0.21	0.24	5.01	7.34E-09	1.40E-05	NCBI_Gene:100861632
TC1300000099.rn.1	15.31	9.82	0.64	0.35	45.09	8.50E-09	1.58E-05	paker; tachee
TC1700001006.rn.1	18.9	14.44	0.36	0.54	21.93	8.92E-09	1.61E-05	jogo
TC0500001378.rn.1	8.3	5.18	0.33	0.32	8.66	9.13E-09	1.61E-05	Cited4
TC0800003273.rn.1	12.68	14.19	0.17	0.14	-2.85	9.45E-09	1.61E-05	RGD1310507
TC1000003889.rn.1	10.29	6.8	0.46	0.17	11.23	9.61E-09	1.61E-05	Kif18b
TC1300001156.rn.1	19.93	15.9	0	0.58	16.36	1.00E-08	1.65E-05	korblaw
TC0700003431.rn.1	8.04	6.39	0.12	0.21	3.14	1.04E-08	1.67E-05	Apol2
TC1900000828.rn.1	19.93	12.16	0	1.07	218.84	1.07E-08	1.68E-05	skador
TC1000004330.rn.1	4.24	7.08	0.18	0.39	-7.16	1.15E-08	1.73E-05	RGD1565158
TC0100002137.rn.1	7.98	5.67	0.3	0.13	4.94	1.18E-08	1.73E-05	P4ha3
TC1200000868.rn.1	7.51	6.32	0.14	0.09	2.29	1.18E-08	1.73E-05	Ulk1
TC0300003734.rn.1	15.47	12.08	0.2	0.49	10.47	1.23E-08	1.77E-05	glarchey
TC0400001871.rn.1	19.93	14.35	0	0.88	48.05	1.29E-08	1.77E-05	spoyzor
TC1700001125.rn.1	19.93	14.35	0	0.88	48.05	1.29E-08	1.77E-05	glarree; spaklu; swergo; wory
TC0100006816.rn.1	9.92	7.52	0.21	0.26	5.28	1.36E-08	1.83E-05	Uros
TC0800002806.rn.1	13.72	16.19	0.14	0.33	-5.52	1.38E-08	1.83E-05	Rab8b
TC0300004938.rn.1	6.22	7.53	0.16	0.07	-2.48	1.46E-08	1.88E-05	ENSMUSG00000087442
TC0600002517.rn.1	7.47	5.1	0.28	0.23	5.16	1.52E-08	1.88E-05	blawtu; LOC102552151
TC1000001359.rn.1	15.39	19.93	0.65	0	-23.33	1.53E-08	1.88E-05	lornaw
TC2000000820.rn.1	19.93	14.14	0	0.92	55.37	1.53E-08	1.88E-05	tado
TC0700001117.rn.1	19.93	14.73	0	0.83	36.86	1.59E-08	1.88E-05	glyser
TC1500002369.rn.1	18.55	15.6	0	0.47	7.71	1.59E-08	1.88E-05	rujer
TC0300002947.rn.1	7.35	6.39	0.09	0.1	1.94	1.61E-08	1.88E-05	Ndufa8
TC0900000051.rn.1	19.93	15.73	0	0.64	18.43	1.64E-08	1.89E-05	fleynoy

Table B.4: Microarray generated output for n=6 differential gene expression analysis of WKY juvenile blood vs SHR juvenile blood at 4 weeks of age. Here, differential expression analysis was conducted on Robust Multiarray Averaging (RMA) normalised microarray intensities using Thermofisher's Transcriptome Analysis Console (TAC, ThermoFisher Scientific). Displayed are the Affymetrics $RTA_1.0$. probe accession IDs, the average Log2 transformed intensities for SHR and WKY samples, the corresponding standard deviations (SD), the fold change from WKY as the control group, the associated P-Value, the multiple test corrected P-Value (FDR), and the Gene name for each transcript. Transcripts have been ranked by False Discovery Rate normalised P-Value (Smallest to largest) to correct for multiple testing, before only including protein-coding genes.

Appendix C

Chapter 5

```
>Ifit1A_MERGE

TACTTAAGCTTGGTACCGAGCTCGGATCCGCCACCATGGGAGAGAATGCTGGTGGTGACCAGGTCATGG
AGAATCTGCTTCAGCTGAGATGTCACTTCACATGGGGCCTGCTCTTTGAAAAAAATGACATACCTGATT
TGGAAGTGAGGATCTCAGAGCAGGTCCAGTTCCTTGACATCAAGAACTCACTGGGGATGCACAACCTCC
AGGCCTACGTGAGACACCTGAAAGGTGAGCAGGAGGAAGCCCTGCAGAGCTTGAAAGAGGCTGAAGCCT
TGATTGAGGGAGAGCAGTTGGGCAAGAGAAGCCTGGTGACCTGGGGCAACTGTGCCTGGGTGCATTACC
ACAGGGGCAGCTTGGCAGAAGCCCAGATCTACCTGGACAAGGTGGAGAATGTTTGCAGGGAGTTCTCAA
GTCCCTTCCGCTACAGGATGGAGTGTGCTGAGATAGACTGTGAGGAAGGCTGGGCCTTGCTGAAGTGTG
GAGGAAGTAACTATATGCGAGCCATGGCCTGCTTTGCAAAGGCTCTGCAGGTGGACCCAGAAAACCCTG
AGTACAATGCTGGCTATGCAGTTGTGGCCTATCGCCAAGATTTCGATGACAACCATGTTTCTCTAGAAC
CTCTAAGGAAGGCTGTCAGGTTAAATCCAGAAGATCCACACCTTAAAGTTCTCCTTGCTCTAAAGCTTCA
GGATTTAGGGGAACAAGATGAAGCAGAAACACACATTGAAGAAGCCCTCAGCAGCACATCTTGCCAAAG
CTATGTCTTTCGCTACGCAGCCAAGTATTTCCGCCGGAAAGGTGACATAAACGAAGCTCTTCATCTTCTA
CACAGGGCCTTGCAGACGTCACCTTCCTCTGGCTACCTACATTACCAAAAAGGGCTCTGTTACAAGCAAC
AGATGATCCAACTGAAGACATCTGGAAACAGGCAGGCCAGAAGGCAGGAGAATATACAGGAATTGGCAC
ATCAGGCCATTTGTGAATTTCAAGAGACTTTGAATCTGAGGCCCACATTTGAGATGGCTTACGTTTGCA
TGGCTGAAATGCAGGCAGAAATTGGCCAATATGAAGAAGCAGAGGGAAATTTCCAGGAGGCCCTGAACC
TCAACAACCTTGTAGCCCACATAGAGCAGGATATTCATTTCCGCTACGGCCGTTTCCAACAATTTCATAA
GAAGTCAGAAGACAAGGCAATCACTCACTACTTAAAAGGTCTAAAACTAGAAGAGAAGTCCTTTGCCTG
GAGGAAACTACTCACAGCTTTGGAGAAAGTGGCTGAAAGACGTGTTCGCCAGAATGTCCGGCTTGTGGA
GAGCACCAGCCTTCTTGGACTGGTCTTCAAACTGAAAGGGCAGGAGATGAAGGCCCTGCTTTACTATGA
GAAGGCCCTGAGGCTCTCTGGGGAAATGAACCCTGCATTTTGAATGCAGCCTCGAGTCTAGAGGCCCC

BLASTN 2.9.0+
Reference: Zheng Zhang, Scott Schwartz, Lukas Wagner, and
Webb Miller (2000), "A greedy algorithm for aligning DNA
sequences", J Comput Biol 2000; 7(1-2):203-14.


Reference for database indexing: Aleksandr Morgulis, George
Coulouris, Yan Raytselis, Thomas L. Madden, Richa Agarwala,
Alejandro A. Schaffer (2008), "Database Indexing for
Production MegaBLAST Searches", Bioinformatics 24:1757-1764.

Job Title: Ifit1A_MERGE
Query: None Query ID: lcl|Query_138329 Length: 1444

Rattus norvegicus interferon-induced protein with tetratricopeptide repeats 1 (Ifit1), mRNA
Sequence ID: NM_020096.1 Length: 2174
Number of Matches: 1
See 1 more title(s):
Rattus norvegicus mRNA for glucocorticoid-attenuated response gene 16 product (garg16 gene)
Sequence ID: AJ276893.1 Length: 2174
```

```
Number of Matches:
 Range 1: 86 to 1490

Score:2562 bits(1387), Expect:0.0,
Identities:1399/1405(99\%),  Gaps:0/1405(0\%), Strand: Plus/Plus

Query  31    GCCACCATGGGAGAGAATGCTGGTGGTGACCAGGTCATGGAGAATCTGCTTCAGCTGAGA  90
             || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  86    GCAACCATGGGAGAGAATGCTGGTGGTGACCAGGTCATGGAGAATCTGCTTCAGCTGAGA  145

Query  91    TGTCACTTCACATGGGGCCTGCTCTTTGaaaaaaaTGACATACCTGATTTGGAAGTGAGG  150
             |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Sbjct  146   TGTCACTTCACATGGGGCCTGCTCTTCGAAAAAAATGACATACCTGATTTGGAAGTGAGG  205

Query  151   ATCTCAGAGCAGGTCCAGTTCCTTGACATCAAGAACTCACTGGGGATGCACAACCTCCAG  210
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  206   ATCTCAGAGCAGGTCCAGTTCCTTGACATCAAGAACTCACTGGGGATGCACAACCTCCAG  265

Query  211   GCCTACGTGAGACACCTGAAAGGTGAGCAGGAGGAAGCCCTGCAGAGCTTGAAAGAGGCT  270
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  266   GCCTACGTGAGACACCTGAAAGGTGAGCAGGAGGAAGCCCTGCAGAGCTTGAAAGAGGCT  325

Query  271   GAAGCCTTGATTGAGGGAGAGCAGTTGGGCAAGAGAAGCCTGGTGACCTGGGGCAACTGT  330
             |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  326   GAAGCCTTGATCGAGGGAGAGCAGTTGGGCAAGAGAAGCCTGGTGACCTGGGGCAACTGT  385

Query  331   GCCTGGGTGCATTACCACAGGGGCAGCTTGGCAGAAGCCCAGATCTACCTGGACAAGGTG  390
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  386   GCCTGGGTGCATTACCACAGGGGCAGCTTGGCAGAAGCCCAGATCTACCTGGACAAGGTG  445

Query  391   GAGAATGTTTGCAGGGAGTTCTCAAGTCCCTTCCGCTACAGGATGGAGTGTGCTGAGATA  450
             |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Sbjct  446   GAGAATGTTTGCAGGGAATTCTCAAGTCCCTTCCGCTACAGGATGGAGTGTGCTGAGATA  505

Query  451   GACTGTGAGGAAGGCTGGGCCTTGCTGAAGTGTGGAGGAAGTAACTATATGCGAGCCATG  510
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  506   GACTGTGAGGAAGGCTGGGCCTTGCTGAAGTGTGGAGGAAGTAACTATATGCGAGCCATG  565

Query  511   GCCTGCTTTGCAAAGGCTCTGCAGGTGGACCCAGAAAACCCTGAGTACAATGCTGGCTAT  570
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  566   GCCTGCTTTGCAAAGGCTCTGCAGGTGGACCCAGAAAACCCTGAGTACAATGCTGGCTAT  625

Query  571   GCAGTTGTGGCCTATCGCCAAGATTTCGATGACAACCATGTTTCTCTAGAACCTCTAAGG  630
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  626   GCAGTTGTGGCCTATCGCCAAGATTTCGATGACAACCATGTTTCTCTAGAACCTCTAAGG  685

Query  631   AAGGCTGTCAGGTTAAATCCAGAAGATCCACACCTTAAAGTTCTCCTTGCTCTAAAGCTT  690
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  686   AAGGCTGTCAGGTTAAATCCAGAAGATCCACACCTTAAAGTTCTCCTTGCTCTAAAGCTT  745

Query  691   CAGGATTTAGGGGAACAAGATGAAGCAGAAACACACATTGAAGAAGCCCTCAGCAGCACA  750
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  746   CAGGATTTAGGGGAACAAGATGAAGCAGAAACACACATTGAAGAAGCCCTCAGCAGCACA  805

Query  751   TCTTGCCAAAGCTATGTCTTTCGCTACGCAGCCAAGTATTTCCGCCGGAAAGGTGACATA  810
             |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Sbjct  806   TCTTGCCAAAGCTATGTCTTTCGCTACGCAGCCAAATATTTCCGCCGGAAAGGTGACATA  865

Query  811   AACGAAGCTCTTCATCTTCTACACAGGGCCTTGCAGACGTCACCTTCCTCTGGCTACCTA  870
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  866   AACGAAGCTCTTCATCTTCTACACAGGGCCTTGCAGACGTCACCTTCCTCTGGCTACCTA  925

Query  871   CATTACCAAAAAGGGCTCTGTTACAAGCAACAGATGATCCAACTGAAGACATCTGGAAAC  930
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  926   CATTACCAAAAAGGGCTCTGTTACAAGCAACAGATGATCCAACTGAAGACATCTGGAAAC  985

Query  931   AGGCAGGCCAGAAGGCAGGAGAATATACAGGAATTGGCACATCAGGCCATTTGTGAATTT  990
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  986   AGGCAGGCCAGAAGGCAGGAGAATATACAGGAATTGGCACATCAGGCCATTTGTGAATTT  1045
```

```
Query  991   CAAGAGACTTTGAATCTGAGGCCCACATTTGAGATGGCTTACGTTTGCATGGCTGAAATG  1050
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1046  CAAGAGACTTTGAATCTGAGGCCCACATTTGAGATGGCTTACGTTTGCATGGCTGAAATG  1105

Query  1051  CAGGCAGAAATTGGCCAATATGAAGAAGCAGAGGGAAATTTCCAGGAGGCCCTGAACCTC  1110
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1106  CAGGCAGAAATTGGCCAATATGAAGAAGCAGAGGGAAATTTCCAGGAGGCCCTGAACCTC  1165

Query  1111  AACAACCTTGTAGCCCACATAGAGCAGGATATTCATTTCCGCTACGGCCGTTTCCAACAA  1170
             |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Sbjct  1166  AACAACCTTGTAGCCCACATAGAGCAGGATATTCATTTCCGCTACGGCCGTTCCCAACAA  1225

Query  1171  TTTCATAAGAAGTCAGAAGACAAGGCAATCACTCACTACTTAAAAGGTCTAAAACTAGAA  1230
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1226  TTTCATAAGAAGTCAGAAGACAAGGCAATCACTCACTACTTAAAAGGTCTAAAACTAGAA  1285

Query  1231  GAGAAGTCCTTTGCCTGGAGGAAACTACTCACAGCTTTGGAGAAAGTGGCTGAAAGACGT  1290
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1286  GAGAAGTCCTTTGCCTGGAGGAAACTACTCACAGCTTTGGAGAAAGTGGCTGAAAGACGT  1345

Query  1291  GTTCGCCAGAATGTCCGGCTTGTGGAGAGCACCAGCCTTCTTGGACTGGTCTTCAAACTG  1350
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1346  GTTCGCCAGAATGTCCGGCTTGTGGAGAGCACCAGCCTTCTTGGACTGGTCTTCAAACTG  1405

Query  1351  AAAGGGCAGGAGATGAAGGCCCTGCTTTACTATGAGAAGGCCCTGAGGCTCTCTGGGGAA  1410
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1406  AAAGGGCAGGAGATGAAGGCCCTGCTTTACTATGAGAAGGCCCTGAGGCTCTCTGGGGAA  1465

Query  1411  ATGAACCCTGCATTTTGAATGCAGC  1435
             |||||||||||||||||||||||||
Sbjct  1466  ATGAACCCTGCATTTTGAATGCAGC  1490
```

```
>Ifit1B_MERGE
```

TAACTTAAGCTTGGTACCGAGCTCGGATCCGCCACCATGGGAGAGAATGCTGGTGGTGACCAGGTCATG
GAGAATCTGCTTCAGCTGAGATGTCACTTCACATGGGGCCTGCTCTTTGAAAAAAATGACATACCTGAT
TTGGAAGTGAGGATCTCAGAGCAGGTCCAGTTCCTTGACATCAAGAACTCACTGGGGATGCACAACCTC
CAGGCCTACGTGAGACACCTGAAAGGTGAGCAGGAGGAAGCCCTGCAGAGCTTGAAAGAGGCTGAAGC
CTTGATTGAGGGAGAGCAGTTGGGCAAGAGAAGCCTGGTGACCTGGGGCAACTGTGCCTGGGTGCATTA
CCACAGGGGCAGCTTGGCAGAAGCCCAGATCTACCTGGACAAGGTGGAGAATGTTTGCAGGGAGTTCTC
AAGTCCCTTCCGCTACAGGATGGAGTGTGCTGAGATAGACTGTGAGGAAGGCTGGGCCTTGCTGAAGTG
TGGAGGAAGTAACTATATGCGAGCCATGGCCTGCTTTGCAAAGGCTCTGCAGGTGGACCCAGAAAACCC
TGAGTACAATGCTGGCTATGCAGTTGTGGCCTATCGCCAAGATTTCGATGACAACCATGTTTCTCTAGA
ACCTCTAAGGAAGGCTGTCAGGTTAAATCCAGAAGATCCACACCTTAAAGTTCTCCTTGCTCTAAAGCTT
CAGGATTTAGGGGAACAAGATGAAGCAGAAACACACATTGAAGAAGCCCTCAGCAGCACATCTTGCCAA
AGCTATGTCTTTCGCTACGCAGCCAAGTATTTCCGCCGGAAAGGTGACATAAACGAAGCTCTTCATCTTC
TACACAGGGCCTTGCAGACGTCACCTTCCTCTGGCTACCTACATTACCAAAAAGGGCTCTGTTACAAGCA
ACAGATGATCCAACTGAAGACATCTGGAAACAGGCAGGCCAGAAGGCAGGAGAATATACAGGAATTGGC
ACATCAGGCCATTTGTGAATTTCAAGAGACTTTGAATCTGAGGCCCACATTTGAGATGGCTTACGTTTG
CATGGCTGAAATGCAGGCAGAAATTGGCCAATATGAAGAAGCAGAGGGAAATTTCCAGGAGGCCCTGAA
CCTCAACAACCTTGTAGCCCACATAGAGCAGGATATTCATTTCCGCTACGGCCGTTTCCAACAATTTCAT
AAGAAGTCAGAAGACAAGGCAATCACTCACTACTTAAAAGGTCTAAAACTAGAAGAGAAGTCCTTTGCC
TGGAGGAAACTACTCACAGCTTTGGAGAAAGTGGCTGAAAGACGTGTTCGCCAGAATGTCCGGCTTGTG
GAGAGCACCAGCCTTCTTGGACTGGTCTTCAAACTGAAAGGGCAGGAGATGAAGGCCCTGCTTTACTAT
GAGAAGGCCCTGAGGCTCTCTGGGGAAATGAACCCTGCATTTTGAATGCAGCCTCGAGTCT

```
Job Title: Ifit1B_MERGE
Query: None Query ID: lcl|Query_241007 Length: 1444

Rattus norvegicus interferon-induced protein with tetratricopeptide repeats 1 (Ifit1), mRNA
Sequence ID: NM_020096.1 Length: 2174
Number of Matches: 1
See 1 more title(s):
Rattus norvegicus mRNA for glucocorticoid-attenuated response gene 16 product (garg16 gene)
Sequence ID: AJ276893.1 Length: 2174
Number of Matches:
 Range 1: 86 to 1490
```

```
Score:2562 bits(1387), Expect:0.0,
Identities:1399/1405(99\%),  Gaps:0/1405(0\%), Strand: Plus/Plus

Query  31    GCCACCATGGGAGAGAATGCTGGTGGTGACCAGGTCATGGAGAATCTGCTTCAGCTGAGA  90
             || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  86    GCAACCATGGGAGAGAATGCTGGTGGTGACCAGGTCATGGAGAATCTGCTTCAGCTGAGA  145

Query  91    TGTCACTTCACATGGGGCCTGCTCTTTGaaaaaaaTGACATACCTGATTTGGAAGTGAGG  150
             |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Sbjct  146   TGTCACTTCACATGGGGCCTGCTCTTCGAAAAAAATGACATACCTGATTTGGAAGTGAGG  205

Query  151   ATCTCAGAGCAGGTCCAGTTCCTTGACATCAAGAACTCACTGGGGATGCACAACCTCCAG  210
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  206   ATCTCAGAGCAGGTCCAGTTCCTTGACATCAAGAACTCACTGGGGATGCACAACCTCCAG  265

Query  211   GCCTACGTGAGACACCTGAAAGGTGAGCAGGAGGAAGCCCTGCAGAGCTTGAAAGAGGCT  270
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  266   GCCTACGTGAGACACCTGAAAGGTGAGCAGGAGGAAGCCCTGCAGAGCTTGAAAGAGGCT  325

Query  271   GAAGCCTTGATTGAGGGAGAGCAGTTGGGCAAGAGAAGCCTGGTGACCTGGGGCAACTGT  330
             |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  326   GAAGCCTTGATCGAGGGAGAGCAGTTGGGCAAGAGAAGCCTGGTGACCTGGGGCAACTGT  385

Query  331   GCCTGGGTGCATTACCACAGGGGCAGCTTGGCAGAAGCCCAGATCTACCTGGACAAGGTG  390
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  386   GCCTGGGTGCATTACCACAGGGGCAGCTTGGCAGAAGCCCAGATCTACCTGGACAAGGTG  445

Query  391   GAGAATGTTTGCAGGGAGTTCTCAAGTCCCTTCCGCTACAGGATGGAGTGTGCTGAGATA  450
             |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Sbjct  446   GAGAATGTTTGCAGGGAATTCTCAAGTCCCTTCCGCTACAGGATGGAGTGTGCTGAGATA  505

Query  451   GACTGTGAGGAAGGCTGGGCCTTGCTGAAGTGTGGAGGAAGTAACTATATGCGAGCCATG  510
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  506   GACTGTGAGGAAGGCTGGGCCTTGCTGAAGTGTGGAGGAAGTAACTATATGCGAGCCATG  565

Query  511   GCCTGCTTTGCAAAGGCTCTGCAGGTGGACCCAGAAAACCCTGAGTACAATGCTGGCTAT  570
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  566   GCCTGCTTTGCAAAGGCTCTGCAGGTGGACCCAGAAAACCCTGAGTACAATGCTGGCTAT  625

Query  571   GCAGTTGTGGCCTATCGCCAAGATTTCGATGACAACCATGTTTCTCTAGAACCTCTAAGG  630
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  626   GCAGTTGTGGCCTATCGCCAAGATTTCGATGACAACCATGTTTCTCTAGAACCTCTAAGG  685

Query  631   AAGGCTGTCAGGTTAAATCCAGAAGATCCACACCTTAAAGTTCTCCTTGCTCTAAAGCTT  690
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  686   AAGGCTGTCAGGTTAAATCCAGAAGATCCACACCTTAAAGTTCTCCTTGCTCTAAAGCTT  745

Query  691   CAGGATTTAGGGGAACAAGATGAAGCAGAAACACACATTGAAGAAGCCCTCAGCAGCACA  750
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  746   CAGGATTTAGGGGAACAAGATGAAGCAGAAACACACATTGAAGAAGCCCTCAGCAGCACA  805

Query  751   TCTTGCCAAAGCTATGTCTTTCGCTACGCAGCCAAGTATTTCCGCCGGAAAGGTGACATA  810
             |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Sbjct  806   TCTTGCCAAAGCTATGTCTTTCGCTACGCAGCCAAATATTTCCGCCGGAAAGGTGACATA  865

Query  811   AACGAAGCTCTTCATCTTCTACACAGGGCCTTGCAGACGTCACCTTCCTCTGGCTACCTA  870
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  866   AACGAAGCTCTTCATCTTCTACACAGGGCCTTGCAGACGTCACCTTCCTCTGGCTACCTA  925

Query  871   CATTACCAAAAAGGGCTCTGTTACAAGCAACAGATGATCCAACTGAAGACATCTGGAAAC  930
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  926   CATTACCAAAAAGGGCTCTGTTACAAGCAACAGATGATCCAACTGAAGACATCTGGAAAC  985

Query  931   AGGCAGGCCAGAAGGCAGGAGAATATACAGGAATTGGCACATCAGGCCATTTGTGAATTT  990
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  986   AGGCAGGCCAGAAGGCAGGAGAATATACAGGAATTGGCACATCAGGCCATTTGTGAATTT  1045

Query  991   CAAGAGACTTTGAATCTGAGGCCCACATTTGAGATGGCTTACGTTTGCATGGCTGAAATG  1050
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1046  CAAGAGACTTTGAATCTGAGGCCCACATTTGAGATGGCTTACGTTTGCATGGCTGAAATG  1105

Query  1051  CAGGCAGAAATTGGCCAATATGAAGAAGCAGAGGGAAATTTCCAGGAGGCCCTGAACCTC  1110
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1106  CAGGCAGAAATTGGCCAATATGAAGAAGCAGAGGGAAATTTCCAGGAGGCCCTGAACCTC  1165

Query  1111  AACAACCTTGTAGCCCACATAGAGCAGGATATTCATTTCCGCTACGGCCGTTTCCAACAA  1170
            |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Sbjct  1166  AACAACCTTGTAGCCCACATAGAGCAGGATATTCATTTCCGCTACGGCCGTTCCCAACAA  1225

Query  1171  TTTCATAAGAAGTCAGAAGACAAGGCAATCACTCACTACTTAAAAGGTCTAAAACTAGAA  1230
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1226  TTTCATAAGAAGTCAGAAGACAAGGCAATCACTCACTACTTAAAAGGTCTAAAACTAGAA  1285

Query  1231  GAGAAGTCCTTTGCCTGGAGGAAACTACTCACAGCTTTGGAGAAAGTGGCTGAAAGACGT  1290
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1286  GAGAAGTCCTTTGCCTGGAGGAAACTACTCACAGCTTTGGAGAAAGTGGCTGAAAGACGT  1345

Query  1291  GTTCGCCAGAATGTCCGGCTTGTGGAGAGCACCAGCCTTCTTGGACTGGTCTTCAAACTG  1350
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1346  GTTCGCCAGAATGTCCGGCTTGTGGAGAGCACCAGCCTTCTTGGACTGGTCTTCAAACTG  1405

Query  1351  AAAGGGCAGGAGATGAAGGCCCTGCTTTACTATGAGAAGGCCCTGAGGCTCTCTGGGGAA  1410
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1406  AAAGGGCAGGAGATGAAGGCCCTGCTTTACTATGAGAAGGCCCTGAGGCTCTCTGGGGAA  1465

Query  1411  ATGAACCCTGCATTTTGAATGCAGC  1435
            |||||||||||||||||||||||||
Sbjct  1466  ATGAACCCTGCATTTTGAATGCAGC  1490
```

>Ifit1C_MERGE

TTACTTAAGCTTGGTACCGAGCTCGGATCCGCCACCATGGGAGAGAATGCTGGTGGTGACCAGGTCATG
GAGAATCTGCTTCAGCTGAGATGTCACTTCACATGGGGCCTGCTCTTTGAAAAAAATGACATACCTGAT
TTGGAAGTGAGGATCTCAGAGCAGGTCCAGTTCCTTGACATCAAGAACTCACTGGGGATGCACAACCTC
CAGGCCTACGTGAGACACCTGAAAGGTGAGCAGGAGGAAGCCCTGCAGAGCTTGAAAGAGGCTGAAGC
CTTGATTGAGGGAGAGCAGTTGGGCAAGAGAAGCCTGGTGACCTGGGGCAACTGTGCCTGGGTGCATTA
CCACAGGGGCAGCTTGGCAGAAGCCCAGATCTACCTGGACAAGGTGGAGAATGTTTGCAGGGAGTTCTC
AAGTCCCTTCCGCTACAGGATGGAGTGTGCTGAGATAGACTGTGAGGAAGGCTGGGCCTTGCTGAAGTG
TGGAGGAAGTAACTATATGCGAGCCATGGCCTGCTTTGCAAAGGCTCTGCAGGTGGACCCAGAAAACCC
TGAGTACAATGCTGGCTATGCAGTTGTGGCCTATCGCCAAGATTTCGATGACAACCATGTTTCTCTAGA
ACCTCTAAGGAAGGCTGTCAGGTTAAATCCAGAAGATCCACACCTTAAAGTTCTCCTTGCTCTAAAGCTT
CAGGATTTAGGGGAACAAGATGAAGCAGAAACACACATTGAAGAAGCCCTCAGCAGCACATCTTGCCAA
AGCTATGTCTTTCGCTACGCAGCCAAGTATTTCCGCCGGAAAGGTGACATAAACGAAGCTCTTCATCTTC
TACACAGGGCCTTGCAGACGTCACCTTCCTCTGGCTACCTACATTACCAAAAAGGGCTCTGTTACAAGCA
ACAGATGATCCAACTGAAGACATCTGGAAACAGGCAGGCCAGAAGGCAGGAGAATATACAGGAATTGGC
ACATCAGGCCATTTGTGAATTTCAAGAGACTTTGAATCTGAGGCCCACATTTGAGATGGCTTACGTTTG
CATGGCTGAAATGCAGGCAGAAATTGGCCAATATGAAGAAGCAGAGGGAAATTTCCAGGAGGCCCTGAA
CCTCAACAACCTTGTAGCCCACATAGAGCAGGATATTCATTTCCGCTACGGCCGTTTCCAACAATTTCAT
AAGAAGTCAGAAGACAAGGCAATCACTCACTACTTAAAAGGTCTAAAACTAGAAGAGAAGTCCTTTGCC
TGGAGGAAACTACTCACAGCTTTGGAGAAAGTGGCTGAAAGACGTGTTCGCCAGAATGTCCGGCTTGTG
GAGAGCACCAGCCTTCTTGGACTGGTCTTCAAACTGAAAGGGCAGGAGATGAAGGCCCTGCTTTACTAT
GAGAAGGCCCTGAGGCTCTCTGGGGAAATGAACCCTGCATTTTGAATGCAGCCTCGAGTCTAGAGT

```
Query= Ifit1C_MERGE
Length=1449

                                                                    Score        E     Max
Sequences producing significant alignments:                        (Bits)     Value  ident

NM_020096.1 Rattus norvegicus interferon-induced protein with ...  2562       0.0    100\%

ALIGNMENTS
>NM_020096.1 Rattus norvegicus interferon-induced protein with tetratricopeptide
repeats 1 (Ifit1), mRNA
 AJ276893.1 Rattus norvegicus mRNA for glucocorticoid-attenuated response
gene 16 product (garg16 gene)
```

```
Length=2174

 Score = 2562 bits (1387),  Expect = 0.0
 Identities = 1399/1405 (99\%), Gaps = 0/1405 (0\%)
 Strand=Plus/Plus

Query  31    GCCACCATGGGAGAGAATGCTGGTGGTGACCAGGTCATGGAGAATCTGCTTCAGCTGAGA  90
             || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  86    GCAACCATGGGAGAGAATGCTGGTGGTGACCAGGTCATGGAGAATCTGCTTCAGCTGAGA  145

Query  91    TGTCACTTCACATGGGGCCTGCTCTTTGaaaaaaaTGACATACCTGATTTGGAAGTGAGG  150
             |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Sbjct  146   TGTCACTTCACATGGGGCCTGCTCTTCGAAAAAAATGACATACCTGATTTGGAAGTGAGG  205

Query  151   ATCTCAGAGCAGGTCCAGTTCCTTGACATCAAGAACTCACTGGGGATGCACAACCTCCAG  210
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  206   ATCTCAGAGCAGGTCCAGTTCCTTGACATCAAGAACTCACTGGGGATGCACAACCTCCAG  265

Query  211   GCCTACGTGAGACACCTGAAAGGTGAGCAGGAGGAAGCCCTGCAGAGCTTGAAAGAGGCT  270
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  266   GCCTACGTGAGACACCTGAAAGGTGAGCAGGAGGAAGCCCTGCAGAGCTTGAAAGAGGCT  325

Query  271   GAAGCCTTGATTGAGGGAGAGCAGTTGGGCAAGAGAAGCCTGGTGACCTGGGGCAACTGT  330
             |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  326   GAAGCCTTGATCGAGGGAGAGCAGTTGGGCAAGAGAAGCCTGGTGACCTGGGGCAACTGT  385

Query  331   GCCTGGGTGCATTACCACAGGGGCAGCTTGGCAGAAGCCCAGATCTACCTGGACAAGGTG  390
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  386   GCCTGGGTGCATTACCACAGGGGCAGCTTGGCAGAAGCCCAGATCTACCTGGACAAGGTG  445

Query  391   GAGAATGTTTGCAGGGAGTTCTCAAGTCCCTTCCGCTACAGGATGGAGTGTGCTGAGATA  450
             |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Sbjct  446   GAGAATGTTTGCAGGGAATTCTCAAGTCCCTTCCGCTACAGGATGGAGTGTGCTGAGATA  505

Query  451   GACTGTGAGGAAGGCTGGGCCTTGCTGAAGTGTGGAGGAAGTAACTATATGCGAGCCATG  510
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  506   GACTGTGAGGAAGGCTGGGCCTTGCTGAAGTGTGGAGGAAGTAACTATATGCGAGCCATG  565

Query  511   GCCTGCTTTGCAAAGGCTCTGCAGGTGGACCCAGAAAACCCTGAGTACAATGCTGGCTAT  570
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  566   GCCTGCTTTGCAAAGGCTCTGCAGGTGGACCCAGAAAACCCTGAGTACAATGCTGGCTAT  625

Query  571   GCAGTTGTGGCCTATCGCCAAGATTTCGATGACAACCATGTTTCTCTAGAACCTCTAAGG  630
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  626   GCAGTTGTGGCCTATCGCCAAGATTTCGATGACAACCATGTTTCTCTAGAACCTCTAAGG  685

Query  631   AAGGCTGTCAGGTTAAATCCAGAAGATCCACACCTTAAAGTTCTCCTTGCTCTAAAGCTT  690
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  686   AAGGCTGTCAGGTTAAATCCAGAAGATCCACACCTTAAAGTTCTCCTTGCTCTAAAGCTT  745

Query  691   CAGGATTTAGGGGAACAAGATGAAGCAGAAACACACATTGAAGAAGCCCTCAGCAGCACA  750
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  746   CAGGATTTAGGGGAACAAGATGAAGCAGAAACACACATTGAAGAAGCCCTCAGCAGCACA  805

Query  751   TCTTGCCAAAGCTATGTCTTTCGCTACGCAGCCAAGTATTTCCGCCGGAAAGGTGACATA  810
             |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Sbjct  806   TCTTGCCAAAGCTATGTCTTTCGCTACGCAGCCAAATATTTCCGCCGGAAAGGTGACATA  865

Query  811   AACGAAGCTCTTCATCTTCTACACAGGGCCTTGCAGACGTCACCTTCCTCTGGCTACCTA  870
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  866   AACGAAGCTCTTCATCTTCTACACAGGGCCTTGCAGACGTCACCTTCCTCTGGCTACCTA  925

Query  871   CATTACCAAAAAGGGCTCTGTTACAAGCAACAGATGATCCAACTGAAGACATCTGGAAAC  930
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  926   CATTACCAAAAAGGGCTCTGTTACAAGCAACAGATGATCCAACTGAAGACATCTGGAAAC  985

Query  931   AGGCAGGCCAGAAGGCAGGAGAATATACAGGAATTGGCACATCAGGCCATTTGTGAATTT  990
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  986   AGGCAGGCCAGAAGGCAGGAGAATATACAGGAATTGGCACATCAGGCCATTTGTGAATTT  1045
```

```
Query  991   CAAGAGACTTTGAATCTGAGGCCCACATTTGAGATGGCTTACGTTTGCATGGCTGAAATG  1050
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1046  CAAGAGACTTTGAATCTGAGGCCCACATTTGAGATGGCTTACGTTTGCATGGCTGAAATG  1105

Query  1051  CAGGCAGAAATTGGCCAATATGAAGAAGCAGAGGGAAATTTCCAGGAGGCCCTGAACCTC  1110
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1106  CAGGCAGAAATTGGCCAATATGAAGAAGCAGAGGGAAATTTCCAGGAGGCCCTGAACCTC  1165

Query  1111  AACAACCTTGTAGCCCACATAGAGCAGGATATTCATTTCCGCTACGGCCGTTTCCAACAA  1170
             |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Sbjct  1166  AACAACCTTGTAGCCCACATAGAGCAGGATATTCATTTCCGCTACGGCCGTTCCCAACAA  1225

Query  1171  TTTCATAAGAAGTCAGAAGACAAGGCAATCACTCACTACTTAAAAGGTCTAAAACTAGAA  1230
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1226  TTTCATAAGAAGTCAGAAGACAAGGCAATCACTCACTACTTAAAAGGTCTAAAACTAGAA  1285

Query  1231  GAGAAGTCCTTTGCCTGGAGGAAACTACTCACAGCTTTGGAGAAAGTGGCTGAAAGACGT  1290
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1286  GAGAAGTCCTTTGCCTGGAGGAAACTACTCACAGCTTTGGAGAAAGTGGCTGAAAGACGT  1345

Query  1291  GTTCGCCAGAATGTCCGGCTTGTGGAGAGCACCAGCCTTCTTGGACTGGTCTTCAAACTG  1350
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1346  GTTCGCCAGAATGTCCGGCTTGTGGAGAGCACCAGCCTTCTTGGACTGGTCTTCAAACTG  1405

Query  1351  AAAGGGCAGGAGATGAAGGCCCTGCTTTACTATGAGAAGGCCCTGAGGCTCTCTGGGGAA  1410
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1406  AAAGGGCAGGAGATGAAGGCCCTGCTTTACTATGAGAAGGCCCTGAGGCTCTCTGGGGAA  1465

Query  1411  ATGAACCCTGCATTTTGAATGCAGC  1435
             |||||||||||||||||||||||||
Sbjct  1466  ATGAACCCTGCATTTTGAATGCAGC  1490
```

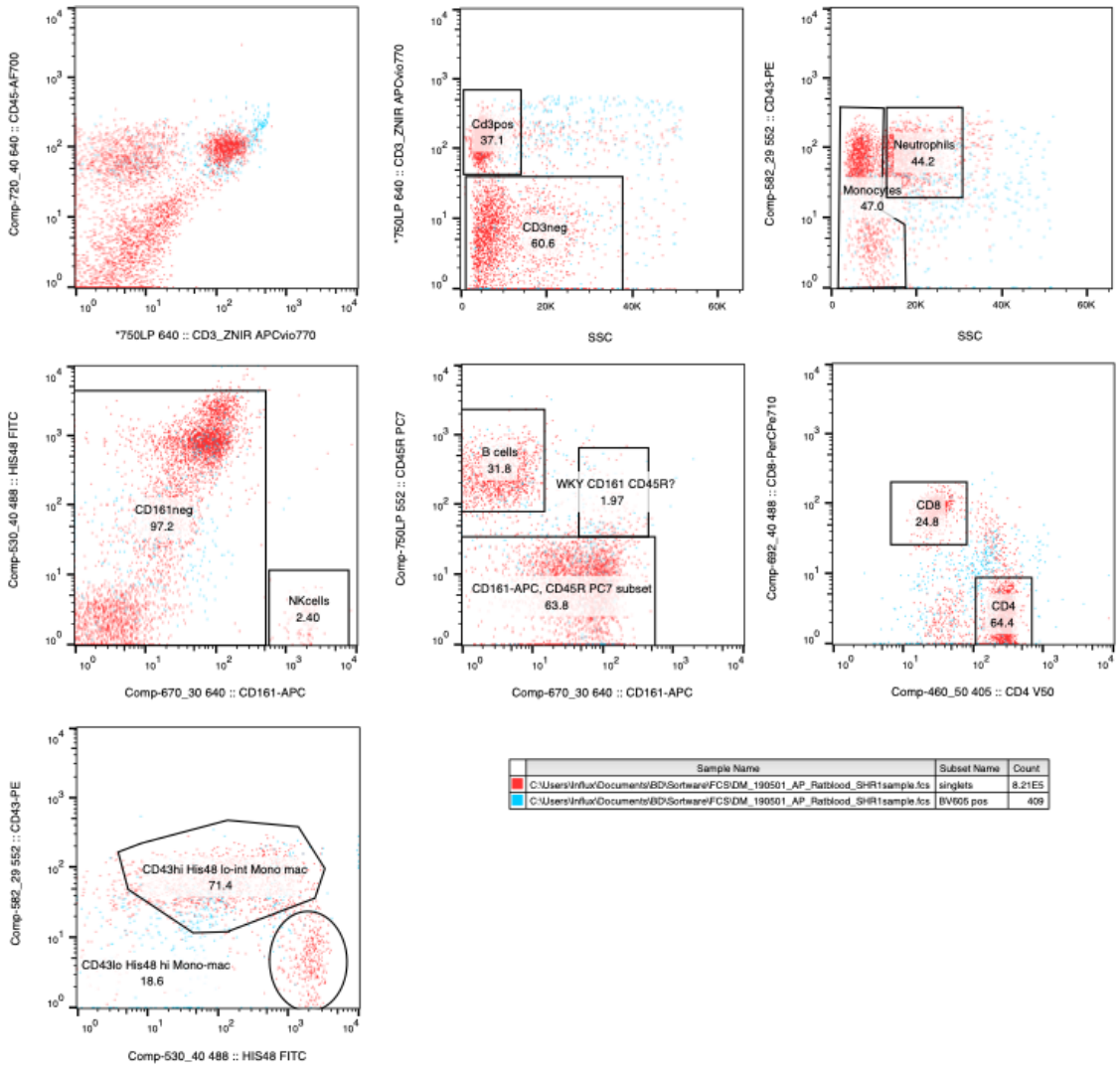


Figure C.1: Example of fluorescence-activated cell sorting (FACS) gating strategy used to isolate Ifit1 expressing cells in SHR blood, following strategy outlined by Barnett-Vanes *et al.*[205] Here, BV605 Positive cells (Stained with anti-Ifit1 antibodies) are highlighted in blue.

Appendix D

Publications

Exp Physiol 102.11 (2017) pp 1373–1379

Experimental Physiology

Symposium Report

Sex-specific differences in cardiovascular and metabolic hormones with integrated signalling in the paraventricular nucleus of the hypothalamus

Spencer P. Loewen[1], Alex R. Paterson[2], Su Yi Loh[3], Mark F. Rogers[2], Charles C. T. Hindmarch[3,4], David Murphy[2,3] and Alastair V. Ferguson[1]

[1]*Centre for Neuroscience, Queen's University, Kingston, Ontario, Canada*
[2]*Henry Wellcome Laboratories for Integrative Neuroscience and Endocrinology, Dorothy Hodgkin Building, University of Bristol, Bristol BS1 3NY, UK*
[3]*Department of Physiology, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia*
[4]*Queen's Cardiopulmonary Unit (QCPU), Translational Institute of Medicine (TIME), Department of Medicine, Queen's University, Kingston, Ontario, Canada*

Edited by: Greti Aguilera

New Findings

- **What is the topic of this review?**
  We describe roles of crucial signalling molecules in the paraventricular nucleus of the hypothalamus and highlight recent data suggesting sex-specific changes in the expression of crucial signalling molecules and their receptors, which may underlie sex differences in both cardiovascular and metabolic function.
- **What advances does it highlight?**
  This review highlights the integrative capacity of the paraventricular nucleus in mediating cardiovascular and metabolic effects by integrating information from multiple signalling molecules. It also proposes that these signalling molecules have sex-specific differential gene expression, indicating the importance of considering these differences in our ongoing search to understand the female–male differences in the regulation of crucial autonomic systems.

Many traditional cardiovascular hormones have been implicated in metabolic function. Conversely, many hormones traditionally involved in metabolic regulation have an effect on cardiovascular function. Many of these signalling molecules exert such effects through specific actions in the paraventricular nucleus, an integrative autonomic control centre located in the hypothalamus. Here, we focus on four cardiovascular/metabolic peptide hormones that signal within the paraventricular nucleus, namely angiotensin II, orexin, adiponectin and nesfatin-1. Each of these hormones has specific electrophysiological effects on paraventricular nucleus neurons that can be related to its physiological actions. In addition, we introduce preliminary transcriptomic data indicating that the genes for some of these hormones and their receptors have sex-specific differential expression.

(Received 6 June 2017; accepted after revision 20 July 2017; first published online 1 August 2017)
Corresponding author A. V. Ferguson: Department of Biomedical and Molecular Sciences, Botterell Hall, Queen's University, Kingston, Ontario, Canada K7L 3N6. Email: avf@queensu.ca

DOI: 10.1113/EP086436

Introduction

The hypothalamic paraventricular nucleus (PVN) is widely known to be a crucial autonomic control centre that plays essential roles in energy homeostasis, fluid balance and cardiovascular regulation, through neuroendocrine and autonomic functions (Swanson & Sawchenko, 1983). Within the PVN exist three subpopulations of neurons, the magnocellular (MNC) neurosecretory, parvocellular pre-autonomic (PA) and parvocellular neuroendocrine (NE) neurons (Swanson & Sawchenko, 1983). The MNC neurons project to the posterior pituitary, where they release vasopressin and oxytocin (OT) into the general circulation. In addition to their neurohypophysial projection, axons of OT-secreting MNC neurons send collaterals to several forebrain regions, including the central amygdala and nucleus accumbens (Knobloch *et al.* 2012). The PA neurons project to autonomic centres in the medulla and spinal cord, including the nucleus tractus solitarii (NTS; Kannan & Yamashita, 1985), to influence autonomic control. Finally, NE neurons project to the median eminence, where they secrete hypophysiotrophic hormones, including corticotrophin-releasing hormone (CRH) and thyrotrophin-releasing hormone.

The PVN receives afferent projections from additional hypothalamic regions, including the lateral hypothalamus and arcuate nucleus (Swanson & Sawchenko, 1983). Inputs are also received from extrahypothalamic sites, including the subfornical organ (Swanson & Sawchenko, 1983), one of the sensory circumventricular organs. The PVN integrates information provided by a vast number of signalling molecules, many of which are circulating hormones that also act as neuropeptides within the CNS (Table 1) and could be thought of collectively as the alphabet for communication in this crucial autonomic control centre. In this review, we highlight four cardiovascular/metabolic peptides that signal within the PVN and present preliminary data indicating that the genes for some of these molecules and their receptors are differentially expressed in males and females.

Table 1. Signalling molecules in the paraventricular nucleus: ABCs of integrated autonomic control?

- Adiponectin
- Adrenomedullin
- Adropin
- Angiotensin
- Atrial natriuretic peptide
- Apelin
- Brain-derived neurotrophic factor
- Bone morphogenetic protein
- Corticotrophin-releasing hormone
- Endothelin
- Estradiol
- Gastrin-releasing peptide
- Ghrelin
- Glucocorticoids
- Glucagon-like peptide 1
- Hydrogen sulfide
- Insulin
- Interleukin
- Kisspeptin
- Leptin
- Macrophage migration inhibitory factor
- Melanocyte-stimulating hormone
- Nesfatin
- Neuronostatin
- Nitric oxide
- Neuropeptide W
- Neuropeptide Y
- Orexin
- Oxytocin
- Pituitary adenylate cyclase-activating polypeptide
- Phoenixin
- Prokineticin 2
- Relaxin
- Substance P
- Vasopressin

Note that references supporting roles for each of these peptides in the paraventricular nucleus can be found through PubMed search for paraventricular nucleus and the relevant peptide. They have not been included here given Journal limitations on the number of references.

Angiotensin II

Angiotensin II (Ang II), a crucial component of the renin–angiotensin system (RAS) and potent vasoconstrictor, is perhaps the most studied peptide hormone involved in cardiovascular regulation. As it is unlikely that Ang II can freely cross the blood–brain barrier, circulating Ang II must signal in the brain through the sensory circumventricular organs. Indeed, a subfornical organ–PVN pathway has been well characterized, with neuronal projections from the subfornical organ to the PVN using Ang II as a neurotransmitter (Li & Ferguson, 1993). The Ang II type 1 (AT_1) receptor is expressed predominantly in NE neurons, specifically in those that express CRH and thyrotrophin-releasing hormone (de Kloet *et al.* 2017). Unsurprisingly, optogenetic stimulation of these neurons results in activation of their respective axes, the hypothalamic–pituitary–adrenal and hypothalamic–pituitary–thyroid (HPT) axes (de Kloet *et al.* 2017). Systemic administration of the AT_1 blocker, losartan, which can cross the blood–brain barrier, abrogates the Ang II-induced hypertensive effects of activation of the subfornical organ–PVN pathway (Li *et al.* 1993).

Angiotensin II has excitatory electrophysiological effects on PVN neurons from all three subpopulations (Fig. 1*A* and *B*). In NE neurons, Ang II directly depolarizes these cells by specifically modulating both a non-selective cationic current and a K^+ current (Latchford & Ferguson, 2005). Angiotensin II has indirect depolarizing effects on MNC neurons mediated via glutamate interneurons (Latchford & Ferguson, 2004) and direct excitatory effects on PA neurons that project to the rostral ventrolateral medulla (Cato & Toney, 2005). With these heterogeneous excitatory effects of Ang II, it is not surprising that this peptide influences more than cardiovascular function. Angiotensin II, along with the other components of the RAS, has also been implicated in energy homeostasis.

Similar to its effects on blood pressure, losartan has also been shown to attenuate body weight gain in obese and

diet-resistant rats fed a high-fat diet (Smith *et al.* 2014). Paraventricular nucleus AT_{1a} knockout mice fed a high-fat diet have increased adiposity attributable to both increased energy consumption and decreased energy expenditure (de Kloet *et al.* 2013). These mice also exhibit decreased blood pressure and hypothalamic mRNA expression of CRH and OT (de Kloet *et al.* 2013). Electrophysiologically, this AT_{1a} deletion eliminates the responsiveness of PVN parvocellular neurons to bath-applied Ang II (de Kloet *et al.* 2013). These results suggest that Ang II signals in the PVN via AT_{1a} receptors in parvocellular neurons to regulate energy balance and to mediate part of its effect on blood pressure.

Preliminary transcriptome data from our laboratory have revealed sex-specific differences between some of the genes for components of the RAS. The gene for the precursor of Ang II, *Agt*, displays the greatest difference in expression, being ∼1.3-fold higher in males *versus* females (Fig. 1*C*). Although the Ang II receptor, AT_1, is not differentially expressed (Fig. 1*C*), the increased expression of AGT may lead to increased levels of Ang II and a subsequent increase in blood pressure. As the RAS is also implicated in energy homeostasis, these sex-specific differences might also contribute to their effects on obesity.

Orexin

Orexins/hypocretins (OX-A and OX-B or hypocretin 1 and hypocretin 2, respectively), as their name suggest, are potent orexigenic peptides derived from a common precursor and are almost exclusively expressed in the lateral hypothalamic area (Sakurai *et al.* 1998). There are two OX receptors, OX_1R and OX_2R, both of which bind each OX isoform (Sakurai *et al.* 1998). OX-A has a higher affinity for the OX_1R than OX-B, whereas both isoforms have equal affinity for the OX_2R (Sakurai *et al.* 1998). Lateral hypothalamic neurons, including those that are OX immunoreactive, project to the PVN (Peyron *et al.*

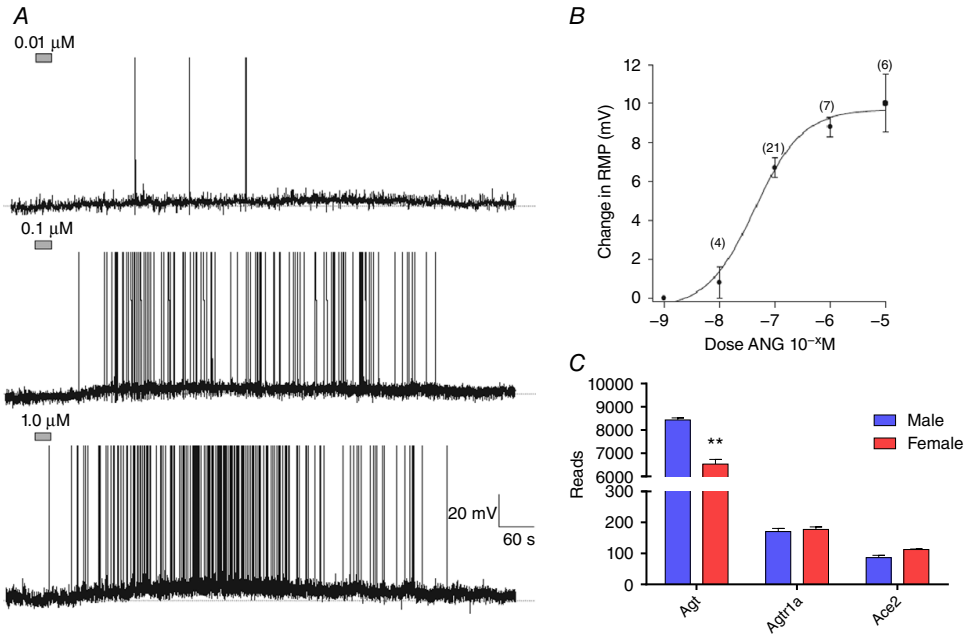


Figure 1. Angiotensin II (Ang II) influences the excitability of neurons in the paraventricular nucleus (PVN)
A, representative current-clamp recordings from three separate PVN neurons exhibiting depolarizations in response to increasing concentrations of bath-applied Ang II (0.01, 0.1 and 1,0 μM, top to bottom). Angiotensin II application is represented by the horizontal grey bars. *B*, mean ± SEM change in resting membrane potential (RMP) elicited by bath application of 10^{-8}, 10^{-7}, 10^{-6} and 10^{-5} M Ang II. Data were fitted to a sigmoid concentration–response function, and the corresponding curve of best fit was overlaid. *C*, preliminary transcriptome data comparing the expression level of angiotensin-related genes in male *versus* female rats. Abbreviations: Ace2, angiotensin-converting enzyme 2; Agt, angiotensinogen; and Agtr1a, angiotensin II type 1a receptor.

1998). Additionally, OX receptors are expressed in PVN neurons (Backberg *et al.* 2002), thus implicating this nucleus in OX-mediated signalling. Indeed, researchers in our laboratory have demonstrated indirect and direct depolarizing effects of OX on MNC and parvocellular PVN neurons, respectively (Fig. 2*A* and *B*; Follwell & Ferguson, 2002).

In addition to their effects on food intake, both OX-A and OX-B have hypertensive effects when administered centrally (Shirasaka *et al.* 1999). In male obese Zucker rats, microinjection of OX-A into the PVN increases blood pressure and renal sympathetic nerve activity, presumably through upregulation of OX_1R in PVN PA neurons projecting to the spinal cord (Zhou *et al.* 2015). These results suggest that OX signalling in the PVN contributes to obesity-related hypertension. In addition, immobilization stress-induced ACTH stimulation is blocked by pretreatment with an OX_1R antagonist, which also abolishes the OX-A-induced increase in blood pressure (Samson *et al.* 2007). These findings suggest that OX-A acts in the PVN to activate the hypothalamic–pituitary–adrenal axis.

Based on our transcriptome data, only one of the orexin receptors, OX_2R, exhibits a sex-specific difference, with greater expression in females than males, whereas the expression of OX_1R is similar in both sexes (Fig. 2*C*). These data are in agreement with a study that looked at nicotine exposure in piglets and observed an increase in OX_2R in the PVN of female animals after nicotine treatment (Hunt *et al.* 2013). These differences may have important implications for the effects of OX in each sex when we consider that the OX isoforms have varying affinities for each OX receptor.

Adiponectin

The adipokine adiponectin (ADP) is involved in glucose and lipid homeostasis. Similar to Ang II, circulating ADP can signal within the brain as a neuropeptide. There are two main ADP receptors (AdipoRs), AdipoR1

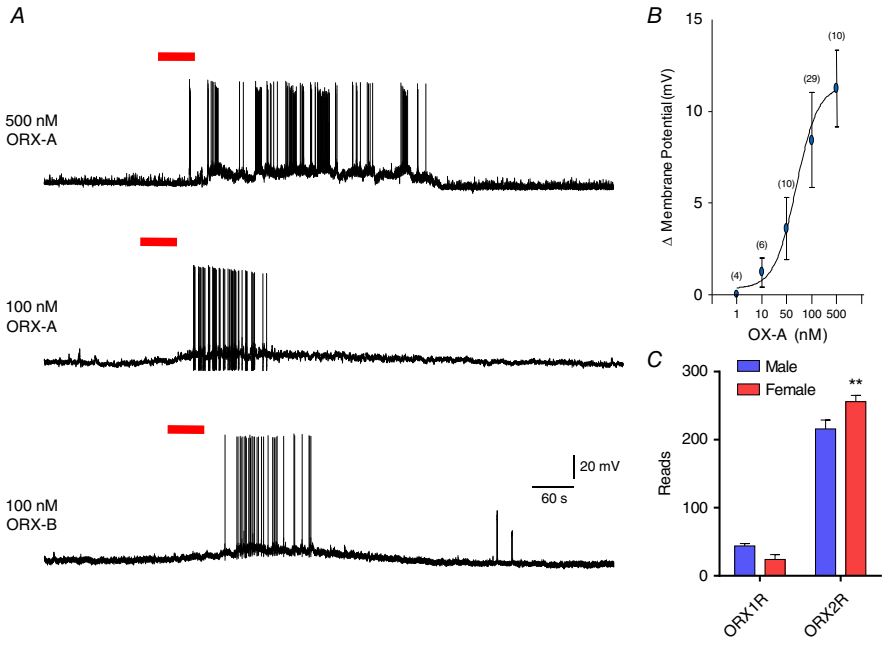


Figure 2. Orexin (OX) influences the excitability of neurons in the PVN
A, representative current-clamp recordings from three separate PVN neurons showing that bath application of OX-A (500 nM, top; 100 nM, middle) or OX-B (100 nM, bottom) results in depolarization of the membrane potential. The time of OX application is indicated by the horizontal red bars. *B*, mean ± SEM change in membrane potential elicited by bath application of 1, 10, 50, 100 and 500 nM OX-A. Data were fitted to a sigmoid concentration–response function, and the corresponding curve of best fit was overlaid. *C*, preliminary transcriptomic data comparing the expression level of the orexin receptor genes in male *versus* female rats. Abbreviations: ORX1R, orexin receptor type 1; and ORX2R, orexin receptor type 2.

and AdipoR2 (Yamauchi *et al.* 2003), both of which are expressed in PVN neurons (Hoyda *et al.* 2007). AdipoR knockout mice exhibit differential effects on energy metabolism. AdipoR1$^{-/-}$ mice have increased adiposity along with decreased glucose tolerance and energy expenditure, whereas AdipoR2$^{-/-}$ mice are resistant to high-fat diet-induced obesity and have improved glucose tolerance (Bjursell *et al.* 2007). Likewise, ADP has differential electrophysiological effects on MNC neurons that appear to be correlated with AdipoR expression, suggesting that neurons expressing both AdipoRs hyperpolarize, whereas those that express only AdipoR2 depolarize (Fig. 3*A*; Hoyda *et al.* 2007). In addition to receptor profiles, researchers in our laboratory have also characterized ADP responses to the molecular phenotype of PVN neurons; MNC neurons that express OT hyperpolarize in response to ADP (Hoyda *et al.* 2007). Intriguingly, this peptide depolarizes PA OT neurons (Hoyda *et al.* 2009*a*), indicating that ADP has opposite effects on functionally distinct OT-expressing PVN neurons.

We have shown a direct depolarizing effect of ADP on NE CRH neurons in the PVN (Fig. 3*B*), leading to subsequent release of ACTH (Hoyda *et al.* 2009*a*). Similar direct depolarizing effects of ADP were found to occur in NTS neurons that express neuropeptide Y, further implicating ADP in feeding behaviour (Hoyda *et al.* 2009*b*). Additionally, direct injection of ADP into the NTS decreases blood pressure (Hoyda *et al.* 2009*b*), and nanoinjection of neuropeptide Y into the PVN induces a similar pressor response (Cassaglia *et al.* 2014). As the parvocellular region of the PVN receives afferent projections from the NTS (Kannan & Yamashita, 1985), these results provide an indirect mechanism of ADP-mediated CRH secretion via the NTS. The reciprocal connections between the PVN and the NTS, and the actions of ADP on both nuclei, suggest a dual effect of this peptide on both feeding behaviour and cardiovascular function.

Although ADP itself does not display a sex-specific difference in expression, one of its receptors does. The gene for AdipoR1 has increased expression in females compared with males (Fig. 3*C*). As ADP signalling depends on the AdipoR expression profile, this differential gene expression might cause differences in the effects of ADP in males *versus* females, at least at the cellular level.

Nesfatin-1

Nesfatin-1 is an anorexigenic peptide derived from the precursor, nucleobindin2 (NUCB2; Oh *et al.* 2006). NUCB2 is expressed in multiple hypothalamic nuclei, including the lateral hypothalamic area, the arcuate nucleus, the supraoptic nucleus and the PVN

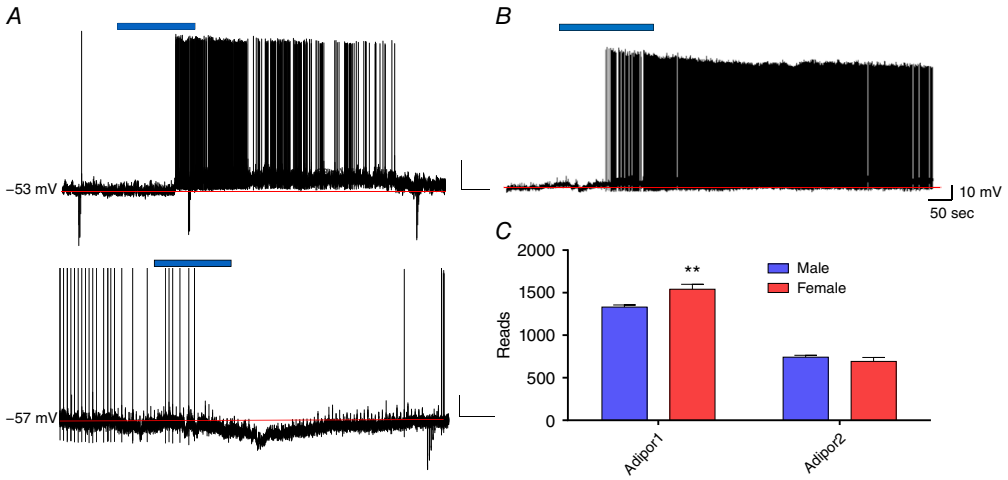


Figure 3. Adiponectin (ADP) influences the excitability of neurons in the PVN
A, representative current-clamp recordings from two separate magnocellular PVN neurons showing that bath application of 10 nM adiponectin (horizontal blue bars) caused a depolarization (top), whereas the second PVN neuron (bottom) exhibited a hyperpolarization. *B*, representative current-clamp recording from a corticotrophin-releasing hormone PVN neuron showing that bath application of 10 nM adiponectin (horizontal blue bar) caused a depolarization. *C*, preliminary transcriptomic data comparing the expression level of the adiponectin receptor genes in male *versus* female rats. Abbreviations: Adipor1, adiponectin receptor 1; and Adipor2, adiponectin receptor 2.

(Oh *et al.* 2006). During starvation, the concentration of nesfatin-1 is reduced in the PVN, while remaining unchanged in the other hypothalamic nuclei (Oh *et al.* 2006). Paraventricular nucleus-specific NUCB2 mRNA knockdown in mice increases food intake, leading to subsequent obesity (Nakata *et al.* 2016). Additionally, central pretreatment with an OT receptor antagonist blocks the inhibitory effect of nesfatin-1 on food intake (Yosten & Samson, 2010), implicating OT in nesfatin-1-mediated feeding behaviour. Clearly, the PVN is involved in mediating the feeding effects of nesfatin-1.

In addition to its role in food intake, nesfatin-1 has also been implicated in cardiovascular function, as central administration has been shown to increase blood pressure in rats (Yosten & Samson, 2009). This nesfatin-1-induced pressor effect occurs through the central melanocortin pathway (Yosten & Samson, 2009), the same pathway used by this peptide for its effect on food intake (Oh *et al.* 2006). Additionally, pretreatment with an OT receptor antagonist abrogates both the anorexigenic and the hypertensive effects of nesfatin-1 (Yosten & Samson, 2010), again implicating OT, and thus the PVN, in mediating these nesfatin-1 effects. Nesfatin-1 has non-specific electrophysiological effects in the PVN, causing both depolarizations and hyperpolarizations in PVN neurons from all three subpopulations (Price *et al.* 2008). Single-cell RT-PCR experiments revealed that these heterogeneous effects are not dependent on the molecular phenotype of these neurons, because both effects occur in PVN neurons that express OT, vasopressin, CRH or thyrotrophin-releasing hormone (Price *et al.* 2008). Additional studies are required to relate the electrophysiological effects of nesfatin-1 in the PVN to a specific physiological action of the peptide; however, it is clear that this nucleus is involved in regulating the cardiovascular and metabolic effects of nesfatin-1. Although the receptor for nesfatin-1 remains elusive, transcriptome data on the precursor gene for nesfatin-1, *Nucb2*, show a similar expression in the PVN of both females and males.

Conclusion

Many signalling molecules that have traditionally been thought to influence only one major system in fact influence multiple systems. These effects, unsurprisingly, are often mediated at least in part by the PVN. We have highlighted four hormones that signal through the PVN and influence both cardiovascular and metabolic function. To study the electrophysiological effects of these hormones within the brain, many studies, including the numerous ones from our laboratory, often obtain data exclusively from male animals, in order to avoid the influence of the female oestrous cycle. However, the preliminary data presented in this review suggest sex-specific differences in gene expression that should be taken into consideration when performing experiments relating to these hormones or their receptors, especially if they involve both male and female animals. Future research should aim at elucidating potential mechanisms underlying these differences and determine whether these differences result in observable effects at the physiological level.

References

Backberg M, Hervieu G, Wilson S & Meister B (2002). Orexin receptor-1 (OX-R1) immunoreactivity in chemically identified neurons of the hypothalamus: focus on orexin targets involved in control of food and water intake. *Eur J Neurosci* **15**, 315–328.

Bjursell M, Ahnmark A, Bohlooly YM, William-Olsson L, Rhedin M, Peng XR, Ploj K, Gerdin AK, Arnerup G, Elmgren A, Berg AL, Oscarsson J & Linden D (2007). Opposing effects of adiponectin receptors 1 and 2 on energy metabolism. *Diabetes* **56**, 583–593.

Cassaglia PA, Shi Z, Li B, Reis WL, Clute-Reinig NM, Stern JE & Brooks VL (2014). Neuropeptide Y acts in the paraventricular nucleus to suppress sympathetic nerve activity and its baroreflex regulation. *J Physiol* **592**, 1655–1675.

Cato MJ & Toney GM (2005). Angiotensin II excites paraventricular nucleus neurons that innervate the rostral ventrolateral medulla: an in vitro patch-clamp study in brain slices. *J Neurophysiol* **93**, 403–413.

de Kloet AD, Pati D, Wang L, Hiller H, Sumners C, Frazier CJ, Seeley RJ, Herman JP, Woods SC & Krause EG (2013). Angiotensin type 1a receptors in the paraventricular nucleus of the hypothalamus protect against diet-induced obesity. *J Neurosci* **33**, 4825–4833.

de Kloet AD, Wang L, Pitra S, Hiller H, Smith JA, Tan Y, Nguyen D, Cahill KM, Sumners C, Stern JE & Krause EG (2017). A unique "angiotensin-sensitive" neuronal population coordinates neuroendocrine, cardiovascular, and behavioral responses to stress. *J Neurosci* **37**, 3478–3490.

Follwell MJ & Ferguson AV (2002). Cellular mechanisms of orexin actions on paraventricular nucleus neurones in rat hypothalamus. *J Physiol* **545**, 855–867.

Hoyda TD, Fry M, Ahima RS & Ferguson AV (2007). Adiponectin selectively inhibits oxytocin neurons of the paraventricular nucleus of the hypothalamus. *J Physiol* **585**, 805–816.

Hoyda TD, Samson WK & Ferguson AV (2009*a*). Adiponectin depolarizes parvocellular paraventricular nucleus neurons controlling neuroendocrine and autonomic function. *Endocrinology* **150**, 832–840.

Hoyda TD, Smith PM & Ferguson AV (2009*b*). Adiponectin acts in the nucleus of the solitary tract to decrease blood pressure by modulating the excitability of neuropeptide Y neurons. *Brain Res* **1256**, 76–84.

Hunt NJ, Waters KA & Machaalani R (2013). Orexin receptors in the developing piglet hypothalamus, and effects of nicotine and intermittent hypercapnic hypoxia exposures. *Brain Res* **1508**, 73–82.

Kannan H & Yamashita H (1985). Connections of neurons in the region of the nucleus tractus solitarius with the hypothalamic paraventricular nucleus: their possible involvement in neural control of the cardiovascular system in rats. *Brain Res* **329**, 205–212.

Knobloch HS, Charlet A, Hoffmann LC, Eliava M, Khrulev S, Cetin AH, Osten P, Schwarz MK, Seeburg PH, Stoop R & Grinevich V (2012). Evoked axonal oxytocin release in the central amygdala attenuates fear response. *Neuron* **73**, 553–566.

Latchford KJ & Ferguson AV (2004). ANG II-induced excitation of paraventricular nucleus magnocellular neurons: a role for glutamate interneurons. *Am J Physiol Regul Integr Comp Physiol* **286**, R894–R902.

Latchford KJ & Ferguson AV (2005). Angiotensin depolarizes parvocellular neurons in paraventricular nucleus through modulation of putative nonselective cationic and potassium conductances. *Am J Physiol Regul Integr Comp Physiol* **289**, R52–R58.

Li Z, Bains JS & Ferguson AV (1993). Functional evidence that the angiotensin antagonist losartan crosses the blood-brain barrier in the rat. *Brain Res Bull* **30**, 33–39.

Li Z & Ferguson AV (1993). Subfornical organ efferents to paraventricular nucleus utilize angiotensin as a neurotransmitter. *Am J Physiol Regul Integr Comp Physiol* **265**, R302–R309.

Nakata M, Gantulga D, Santoso P, Zhang B, Masuda C, Mori M, Okada T & Yada T (2016). Paraventricular NUCB2/Nesfatin-1 supports oxytocin and vasopressin neurons to control feeding behavior and fluid balance in male mice. *Endocrinology* **157**, 2322–2332.

Oh IS, Shimizu H, Satoh T, Okada S, Adachi S, Inoue K, Eguchi H, Yamamoto M, Imaki T, Hashimoto K, Tsuchiya T, Monden T, Horiguchi K, Yamada M & Mori M (2006). Identification of nesfatin-1 as a satiety molecule in the hypothalamus. *Nature* **443**, 709–712.

Peyron C, Tighe DK, van den Pol AN, de Lecea L, Heller HC, Sutcliffe JG & Kilduff TS (1998). Neurons containing hypocretin (orexin) project to multiple neuronal systems. *J Neurosci* **18**, 9996–10015.

Price CJ, Hoyda TD, Samson WK & Ferguson AV (2008). Nesfatin-1 influences the excitability of paraventricular nucleus neurones. *J Neuroendocrinol* **20**, 245–250.

Sakurai T, Amemiya A, Ishii M, Matsuzaki I, Chemelli RM, Tanaka H, Williams SC, Richardson JA, Kozlowski GP, Wilson S, Arch JR, Buckingham RE, Haynes AC, Carr SA, Annan RS, McNulty DE, Liu WS, Terrett JA, Elshourbagy NA, Bergsma DJ & Yanagisawa M (1998). Orexins and orexin receptors: a family of hypothalamic neuropeptides and G protein-coupled receptors that regulate feeding behavior. *Cell* **92**, 573–585.

Samson WK, Bagley SL, Ferguson AV & White MM (2007). Hypocretin/orexin type 1 receptor in brain: role in cardiovascular control and the neuroendocrine response to immobilization stress. *Am J Physiol Regul Integr Comp Physiol* **292**, R382–R387.

Shirasaka T, Nakazato M, Matsukura S, Takasaki M & Kannan H (1999). Sympathetic and cardiovascular actions of orexins in conscious rats. *Am J Physiol Regul Integr Comp Physiol* **277**, R1780–R1785.

Smith PM, Hindmarch CCT, Murphy D & Ferguson AV (2014). AT_1 receptor blockade alters nutritional and biometric development in obesity-resistant and obesity-prone rats submitted to a high fat diet. *Front Psychol* **5**, 832.

Swanson LW & Sawchenko PE (1983). Hypothalamic integration: organization of the paraventricular and supraoptic nuclei. *Annu Rev Neurosci* **6**, 269–324.

Yamauchi T, Kamon J, Ito Y, Tsuchida A, Yokomizo T, Kita S, Sugiyama T, Miyagishi M, Hara K, Tsunoda M, Murakami K, Ohteki T, Uchida S, Takekawa S, Waki H, Tsuno NH, Shibata Y, Terauchi Y, Froguel P, Tobe K, Koyasu S, Taira K, Kitamura T, Shimizu T, Nagai R & Kadowaki T (2003). Cloning of adiponectin receptors that mediate antidiabetic metabolic effects. *Nature* **423**, 762–769.

Yosten GL & Samson WK (2009). Nesfatin-1 exerts cardiovascular actions in brain: possible interaction with the central melanocortin system. *Am J Physiol Regul Integr Comp Physiol* **297**, R330–R336.

Yosten GL & Samson WK (2010). The anorexigenic and hypertensive effects of nesfatin-1 are reversed by pretreatment with an oxytocin receptor antagonist. *Am J Physiol Regul Integr Comp Physiol* **298**, R1642–R1647.

Zhou JJ, Yuan F, Zhang Y & Li DP (2015). Upregulation of orexin receptor in paraventricular nucleus promotes sympathetic outflow in obese Zucker rats. *Neuropharmacology* **99**, 481–490.

Additional information

Competing interests

None declared.

Author contributions

S.P.L., A.R.P., S.Y.L., M.F.R., C.C.T.H., D.M. and A.V.F. all contributed to the writing of this review, approved the final version of the manuscript and agree to be accountable for all aspects of this work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All persons designated as authors qualify for authorship, and all those who qualify for authorship are listed.

Funding

This work was supported by the Canadian Institutes of Health Research (grant MOP12192 to A.V.F.).

Neurobiology of Aging 65 (2018) 178–191

Contents lists available at ScienceDirect

Neurobiology of Aging

journal homepage: www.elsevier.com/locate/neuaging

The effects of aging on biosynthetic processes in the rat hypothalamic osmoregulatory neuroendocrine system

Michael P. Greenwood [a,*,1], Mingkwan Greenwood [a,1], Elena V. Romanova [b], Andre S. Mecawi [c,d,e], Alex Paterson [a], Olivera Sarenac [a,f], Nina Japundžić-Žigon [f], José Antunes-Rodrigues [c], Julian F.R. Paton [g], Jonathan V. Sweedler [b], David Murphy [a,d]

[a] School of Clinical Sciences, University of Bristol, Bristol, England
[b] Department of Chemistry and the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[c] School of Medicine of Ribeirão Preto, University of São Paulo, Ribeirão Preto, Brazil
[d] Department of Physiology, University of Malaya, Kuala Lumpur, Malaysia
[e] Department of Physiological Sciences, Institute of Biological and Health Sciênces, Federal Rural University of Rio de Janeiro, Seropedica, Brazil
[f] Institute of Pharmacology, Clinical Pharmacology and Toxicology, Faculty of Medicine, University of Belgrade, Belgrade, Serbia
[g] School of Physiology and Pharmacology, University of Bristol, Bristol, England

ARTICLE INFO

Article history:
Received 15 August 2016
Received in revised form 11 July 2017
Accepted 16 January 2018
Available online 31 January 2018

Keywords:
Aging
Vasopressin
Supraoptic nucleus
Methylation
Gene expression
Peptidomics

ABSTRACT

Elderly people exhibit a diminished capacity to cope with osmotic challenges such as dehydration. We have undertaken a detailed molecular analysis of arginine vasopressin (AVP) biosynthetic processes in the supraoptic nucleus (SON) of the hypothalamus and secretory activity in the posterior pituitary of adult (3 months) and aged (18 months) rats, to provide a comprehensive analysis of age-associated changes to the AVP system. By matrix-assisted laser desorption/ionization time-of-flight mass spectrometry analysis, we identified differences in pituitary peptides, including AVP, in adult and aged rats under both basal and dehydrated states. In the SON, increased Avp gene transcription, coincided with reduced Avp promoter methylation in aged rats. Based on transcriptome data, we have previously characterized a number of novel dehydration-induced regulatory factors involved in the response of the SON to osmotic cues. We found that some of these increase in expression with age, while dehydration-induced expression of these genes in the SON was attenuated in aged rats. In summary, we show that aging alters the rat AVP system at the genome, transcriptome, and peptidome levels. These alterations however did not affect circulating levels of AVP in basal or dehydrated states.

1. Introduction

As we age, disorders of body salt and water composition become more commonplace. Cases of hyponatremia/hypernatremia are much more prevalent in the elderly, where they have been linked to increased incidences of falls, fractures, and osteoporosis, thus contributing to increased hospital admissions and morbidity and mortality (Cowen et al., 2013). To promote healthy living well into old age, it is thus necessary to determine why such imbalances occur. Age-associated changes to both peripheral and central mechanisms that control salt and water homeostasis are deemed responsible. There is a progressive age-related decline in renal function, with less urine concentrating capacities in the elderly compared with younger subjects (Ishunina and Swaab, 2002). Such impaired capacity to conserve body water, together with reports of reduced thirst and inadequate fluid intake after periods of fluid deprivation, makes the elderly more susceptible to dehydration (Mack et al., 1994; Phillips et al., 1993). Inappropriate release of the antidiuretic hormone arginine vasopressin (AVP) into the systemic circulation has been highlighted as one of the causes of irregular water homeostasis in aging (Swaab and Bao, 2011).

AVP is synthesized in magnocellular neurons of the supraoptic nucleus (SON) and paraventricular nucleus (PVN) of the hypothalamus. A change in plasma osmolality is detected by osmosensitive neurons in circumventricular organs of the brain that provide direct inputs to shape the firing of AVP magnocellular neurons that are osmosensitive themselves and to coordinate AVP synthesis and secretion from the posterior (neural) lobe of the pituitary gland (Mecawi Ade et al., 2015; Nissen et al., 1994; Zhang and Bourque,

* Corresponding author at: University of Bristol, Dorothy Hodgkin Building, Whitson Street, Bristol, BS1 3NY, England. Tel.: +44 (0) 117 3313 071; fax: +44 (0) 117 3313 035.
E-mail address: mike.greenwood@bristol.ac.uk (M.P. Greenwood).
[1] M.P.G. and M.G. contributed equally to this work.

https://doi.org/10.1016/j.neurobiolaging.2018.01.008

2003). Once released, following incidences such as rise in plasma osmolality or decrease in blood volume (Kondo et al., 2004), AVP promotes sodium and water reabsorption by the kidney (Ares et al., 2011; Breyer and Ando, 1994). When placed under stress, the capabilities of the AVP system have been shown to decrease with age (Frolkis et al., 1999; Keck et al., 2000; Sladek et al., 1981).

The AVP system has been interrogated on multiple levels, from synthesis to secretion, in aged subjects with differing results. For example, basal circulating AVP levels have been found to decrease, to remain unchanged, and to increase with age in humans, as well as rodents (Frolkis et al., 1999). These discrepancies have been attributed to genetic, age, and strain differences. Although there are many disputations, 1 area of agreement is that, while in many brain areas neuronal activity decreases with age (Burke and Barnes, 2006), paradoxically AVP neurons become more active (Palin et al., 2009; Terwel et al., 1992). This hyperactivity is thought to be a compensatory mechanism for decreased responsiveness to AVP in the kidney due to decreased receptor abundance in aged subjects (reviewed by Ishunina and Swaab, 2002), although this theory has been questioned (Preisser et al., 2004) and is not intuitive with the profound difference in circulating AVP described in aging models. The AVP magnocellular neurons undergo numerous morphological changes as they age including increased size of perikarya, nucleoli, and Golgi apparatus in humans and rodents (Ishunina and Swaab, 2002), analogous to morphological changes in these neurons with dehydration (Hatton and Walters, 1973). Increased AVP neuron size in states of dehydration is recognized as a necessary measure to meet cellular demands for increased transcription and protein synthesis under hypertonic stimulation where circulating levels of AVP are robustly increased (Zhang et al., 2001). It has been suggested that such hyperactivity of AVP neurons may in itself lead to electrolyte disorders in the elderly (Swaab and Bao, 2011), but the relationship between the activity of AVP neurons and circulating levels of AVP is poorly understood.

Transcriptional changes have been identified in AVP neurons with aging. A study by Palin et al. (Palin et al., 2009) showed increased expression of immediate early gene c-Fos, a commonly used marker of neuronal activity, in the rat SON, consistent with hyperactivation of the AVP neurons. In contrast to increased activity under basal conditions, reports have described an attenuation of the evoked AVP secretion in response to osmotic stress with aging in rodents (Sladek and Olschowka, 1994; Swenson et al., 1997). This has led some to suggest that deficits in mechanisms controlling transcription, mRNA stability, or translation in the aging SON magnocellular neurons may be responsible (Lucassen et al., 1997). Moreover, we recently showed that dehydration initiates the formation of new methylation marks on the rat Avp promoter (Greenwood et al., 2016a), suggesting that altered methylation patterns could lie beneath these transcriptional changes in aging AVP neurons.

Few studies have sought to combine information on the physiological aspects of aging with analyses of the molecular changes occurring in the hypothalamus. The reasons why old AVP neurons have elevated basal activity, or why they can fail to adequately respond under stress, is not well understood. In particular, the molecular basis for these changes in relation to circulating levels of AVP has received little attention. We reasoned that aging-associated deficits in the AVP system may be due to a combination of changes at the genome, transcriptome, and peptidome levels, and that these changes might be responsible for disturbances in osmotic stability. In this study, we have interrogated the physiological aspects of aging, performed metabolic measurements, and analyzed peptide levels in plasma and pituitary, correlating the results with molecular events occurring within the hypothalamus of aged rats. Furthermore, we have used our extensive knowledge of the transcriptome of the adult rat SON in euhydrated and dehydrated states to uncover novel changes surrounding altered Avp transcription in aging.

2. Materials and methods

2.1. Animals

All experiments were performed under a Home Office UK licence held under, and in strict accordance with, the provisions of the UK Animals (Scientific Procedures) Act (1986); they had also been approved by the University of Bristol Animal Welfare and Ethical Review Board. We choose to use male Wistar Han rats from the international genetic standardization program (IGS) in our aging study (Charles River, France). The carefully managed breeding program for these animals helps to manage genetic drift, so colonies bred in different locations around the world are not significantly divergent from each other giving a level continuity in aging studies performed in laboratories worldwide. The Charles River Han Wistar rats have been extensively studied at 2 years of age when these rats are reaching the end of their natural life spans. Incidences of neoplastic and non-neoplastic lesions were high in tissues including the kidney and pituitary gland at this age. Furthermore, rats surviving to this age varied from 30% to 80% across 20 control studies (www.criver.com). Therefore, in this aging study, we opted for rats of 18 months of age to minimize pathophysiological effects and thus allow investigation of the aging process in healthy animals. All adult rats used in this study were free of pituitary tumors; however, 6 of 50 aged animals were removed from this study because of tumors on the pituitary gland. On arrival, rats were 2 weeks younger than the desired ages, 3 months (adult) and 18 months (aged), to enable sufficient time for acclimatization before experimentation. Rats were housed at a constant temperature of 22 °C and a relative humidity of 50%–60% (v/v) under a 14:10-hour light/dark cycle (lights on at 0500) with food and water *ad libitum* for 2 weeks. To induce hyperosmotic stress, both adult and aged rats were randomly assigned to 2 groups: control (free access to drinking water) and dehydrated (removal of drinking water for 3 days). All rats were humanely killed by striking of the cranium (stunning) and then immediately decapitated with a small-animal guillotine (Harvard Apparatus, Holliston, MA). Trunk blood was collected in heparin-coated tubes. Brains were rapidly removed from the cranium and immediately frozen by covering with powdered dry ice (within 3 minutes of stunning). The pituitary gland was removed from the base of the skull within 2 minutes after decapitation. The neurointermediate lobe (NIL) was carefully separated from the anterior pituitary using a scalpel blade and then either placed into 1.5-mL tubes containing 500 μL of 0.1 M of HCl or 0.2-mL tubes containing 150 μL of 15 mg/mL of 2, 5-dihydroxybenzoic acid (DHBA) solution. Frozen brains and NIL in HCl solution were stored at −80 °C, whereas NILs in DHBA solution were stored at 4 °C. Animal experiments were performed between 9 AM and 2 PM.

2.2. Metabolic measures in adult and aged rats

For metabolic measurements, animals were individually housed in metabolic cages (Techniplast, Italy) to allow precise daily measures of fluid and food intake and urine output. A plastic gnawing disc was suspended from the lid of the cage to provide environmental enrichment throughout the study. Animals were first acclimatized to metabolic cages for 48 hours. Measures of food hoppers, water bottles, and urine collection tubes were performed for 3 consecutive days, by weight. Extrarenal secretion of fluid was calculated by subtracting water intake from urine output. Plasma

and urine osmolalities were measured by freezing point depression using a Roebling micro-osmometer (Camlab).

2.3. Vasopressin measures

The NIL of the pituitary was sonicated for 15 seconds in 0.1 M of HCl and incubated at 85 °C for 20 minutes. Cellular debris was removed by centrifugation at 3000 × *g* for 30 minutes at 4 °C. Pituitary AVP content was determined using an AVP[8]-Vasopressin ELISA (ADI-900-017A; Enzo) kit. The supernatant was diluted (1:10,000) with assay buffer, and ELISA was performed following the manufacturer's protocol. The signal was detected on an iMark microplate absorbance reader (Bio-Rad). For radioimmunoassay, trunk blood was centrifuged at 1600 × *g* for 15 minutes at 4 °C. Extractions were performed from 1 mL of plasma. Two sample volumes of ice-cold acetone were added, and samples were vortexed for 1 minute. Protein precipitates were removed by centrifugation at 2500 × *g*, 4 °C, for 25 minutes. The supernatant was transferred to a new tube and mixed with 2 mL of cold petroleum ether by vortexing for 1 minute. The tubes were left to stand for 1 minute at room temperature before discarding the upper phase. The lower phase solution was lyophilized using a freeze dryer (Benchtop Pro; Biopharma). AVP concentration was determined by a specific radioimmunoassay (Husain et al., 1973).

2.4. Peptide analysis of the NIL

Peptides were measured directly in individual NIL extracts by mass spectrometry (MS) (Romanova et al., 2014; Romanova and Sweedler, 2015).

2.4.1. Extraction of peptides

NIL samples were incubated in 15 mg/mL of DHBA solution for 48 hours as described by Romanova et al. (2008). The samples were grouped as follows: adult control (n = 14), adult 3-day dehydrated (n = 16), aged control (n = 11), and aged 3-day dehydrated (n = 13).

2.4.2. Measurement of the NIL peptide profiles by MALDI-TOF MS

For MALDI-TOF MS measurements, 0.7 μL of the NIL extraction solution was spotted on a stainless steel MALDI target in triplicates and cocrystallized with 0.7 μL of a freshly prepared concentrated DHBA matrix (50 mg/mL, 50% (v/v) acetone). Positive ion mass spectra of each spotted sample were acquired automatically at 1-KHz laser frequency, and constant power optimized for the sample type in the 600–6000 *m/z* region using an ultrafleXtreme mass spectrometer (Bruker Daltonics) operated in reflectron mode via the AutoXecute protocol. Acquisition parameters included laser fuzzy control logic, random laser walk over the entire sample area, 250 laser shots per raster step, maximum of 5000 shots per sample in 250-shot increments, and dynamic termination of spectrum acquisition when the signal intensity reached 30,000 counts for 3 peaks, regardless of the number of fired laser shots. Peak evaluation was set to a signal intensity per shot of 20 or above, minimal resolution of 10,000, signal-to-noise ratio (S/N) = 3, maximum 300 peaks per spectrum, and centroid peak detection algorithm; 50 failed spectrum judgments were required before acquisition moving to the next sample. External quadratic calibration was adjusted automatically for every 5 × 5 sample spot square.

2.4.3. Principal component analysis of the peptide profiles

Statistical analysis of raw MALDI MS data was performed using ClinProTools 2.2 software (Bruker Daltonics). All spectra were normalized to total ion count upon loading into ClinProTools and level scaled. Spectra were processed for convex hull baseline correction within 800–5500 *m/z*, smoothed with 0.1 Da × 2 cycles of the Savitzky-Golay method and a data reduction factor of 2, null spectra exclusion was enabled, and spectra grouping applied. Other criteria included automatic peaks selection on the total average group spectrum by intensity, S/N = 5 cut off, 1% relative threshold base peak on average group spectrum, unlimited picking. Manual peak editing for the integration area after automatic peak picking was done on the mean spectrum representative of each sample group to include entire isotopic clusters of highly resolved peaks. Peptide profiles of the mean spectra were compared by principal component analysis (PCA) followed by the Anderson-Darling normality test and *Student's unpaired t*-test for normal distributed data. Data not showing normal distributions ($p_{\text{Anderson-Darling}} \leq 0.05$) were evaluated by Kruskal-Wallis tests, respectively (Kruskal and Wallis, 1952; Stephens, 1974; Wilcoxon, 1945). To decrease the number of false positives while computing individual peak statistics on the complex spectra, the Benjamini-Hochberg procedure incorporated into ClinProTools was automatically applied for *p*-value adjustment during analysis (Dudoit and Shaffer, 2003). Unsupervised clustering of spectra was performed on PCA-modified data using Euclidean distance, average distance methods, and a Minkowski exponent of 1.5. The following peptide profile differences were investigated: (1) between control aged and adult rats; (2) between dehydrated aged and adult rats; and (3) between control and dehydrated rats of either age.

2.5. Dual DNA and RNA extraction from SON punch samples

SON samples (12 unilateral punches) were collected from 12 coronal slices using a 0.35-mm sample corer (Fine Scientific Tools) using the optic chiasm as a reference. Total RNA and genomic DNA were extracted from each sample as previously described (Greenwood et al., 2016a).

2.6. Complementary DNA synthesis and quantitative PCR

For cDNA synthesis, 40 ng of total RNA was reverse transcribed using the QuantiTect reverse transcription (RT) kit (Qiagen). Primers for rat genes used in this study: Avp (5′-TGCCTGCT ACTTCCAGAACTGC-3′ and 5′-AGGGGAGACACTGTCTCAGCTC-3′), heteronuclear Avp (hnAvp) (5′-GAGGCAAGAGGGCCACATC-3′ and 5′-CTCTCCTAGCCCATGACCCTT-3′), ras-related dexamethasone induced 1 (Rasd1) (5′-CCCTCAGCGTTGTGCCTACT-3′ and 5′-AAAGA GCGCACGGAACATCT-3′), caprin family member 2 (Caprin2) (5′-CAG GGTTAAGTGCAAGCGAT-3′ and 5′-CTGGTGGTTGACTGGTTGAG-3′), c-Fos (5′-AGCATGGGCTCCCCTGTCA-3′ and 5′-GAGACCAGAGTGGG CTGCA-3′), cAMP responsive element binding protein 3 like 1 (Creb3l1) (5′-GCCAACAGGACCCTGCTCCA-3′ and 5′-AGTGCCAGT CTGTGTGGCCG-3′), gonadotropin inducible ovarian transcription factor 1 (Giot1) (5′-GACACTTCCGGTCCGTCATAG-3′ and 5′-GCCTCA CTCAAGCACCCAGT-3′), DNA methyltransferase 1 (Dnmt1) (5′-AACC ACTCAGCATTCCCGTA-3′ and 5′-TGCTGGTACTTCAGGTCAGG-3′), Dnmt3a (5′-AAGACCCCTGGAACTGCTAC-3′ and 5′-TGGCGAAGAAC ATCTGGAGT-3′), mature oxytocin (Ot) (5′-TGCCCCAGTCTTGCT TGCT-3′ and 5′-TCCAGGTCTAGCGCAGCCC-3′), heteronuclear Ot (hnOt) (5′-TGAGCAGGAGGGGGCCTAGC-3′ and 5′-TGCAAGAGAAA TGGGTCAGTGGC-3′), proprotein convertase subtilisin/kexin type 1 inhibitor (proSAAS) (5′-GAGCTGCTGAGGTACTTGCT-3′ and 5′-ACC CAAATCCTGGTCCACAG-3′), heteronuclear proSAAS (hnproSAAS) (5′-GAAGTGACGACCGAGGTGTA-3′ and 5′-GCAGTATTGTAGGGCGT TCG-3′), tet methylcytosine dioxygenase 1 (Tet1) (5′-TGACCCACTCT TACCAGACC-3′ and 5′-GATGGGCCATTGCTTGATGT-3′), Tet2 (5′-TCG GAGGAGAAGAGTCAGGA-3′ and 5′-TAGGGCTTGCATTTTCCATC-3′), Tet3 (5′-ATGGCATGAAACCACCCAAC-3′ and 5′-ACTTGATCTTCC CCTCCAGC-3′), and ribosomal protein L19 (Rpl19) (5′-GCGTCT GCAGCCATGAGTA-3′ and 5′-TGGCATTGGCGATTTCGTTG-3′) were

synthesized by Eurofins MWG Operon. QuantiTect Primer Assays for solute carrier family 12, member 1 (Slc12a1) were purchased from Qiagen. The optimization and validation of primers were performed using standard Applied Biosystems protocols. The cDNA from RT reaction was diluted 1:4 with H_2O and used as a template for subsequent polymerase chain reactions (PCRs), which were carried out in duplicate using SYBR green (Roche) on an Applied Biosystems StepOnePlus Real-Time PCR system. For relative quantification of gene expression, the $2^{-\Delta\Delta CT}$ method was used (Livak and Schmittgen, 2001). The internal control gene used for these analyses was the housekeeping gene Rpl19.

2.7. Poly(A) tail–length assay

The poly(A) tail length of the Avp mRNA was examined using the USB poly(A) Tail–Length Assay Kit (Affymetrix). RNA extracted from SON (50 ng) was used as the starting material. Guanosine and inosine residues were added to the 3′ ends of poly(A)-containing RNAs using the poly(A) polymerase enzyme. After incubation at 37 °C for 1 hour, stop solution was added and the tailed-RNAs were converted to cDNA by RT using the newly added G/I tails as priming sites. PCR amplification products were generated by using 2 primer sets: set 1, gene-specific forward and reverse primer set for Avp (forward 5′-CGAGTGTCGAGAGGGTTTTT-3′, reverse 5′-TTTATTTTC CATGCTGTAGG-3′), and set 2, Avp gene–specific forward primer and a universal reverse primer provided in the kit. PCR reactions were performed using 2 μL of the undiluted RT sample. PCRs were performed using the following cycling conditions: 94 °C for 2 minutes followed by 40 cycles of 94 °C for 10 seconds, 60 °C for 45 seconds, and 72 °C for 5 minutes. The PCR products were separated on 2.5% (w/v) agarose/TAE gel. The PCR products were visualized on ethidium bromide–stained gels using a Syngene G:BOX imaging system.

2.8. Bisulfite conversion and sequencing

Genomic DNA from SON punches (25 ng) was bisulfite converted using an EZ DNA Methylation-Gold kit (Zymo Research). The amplification and sequencing steps were performed as previously described (Greenwood et al., 2016a).

3. Results

3.1. Physiological assessment of adult and aged rats

We singly housed adult and aged rats in metabolic cages to assess their ingestive behaviors. As expected, the average weight of aged rats was significantly higher than that of adult rats (Fig. 1A). Despite their larger size, aged rats consumed significantly less food (Fig. 1B) and water (Fig. 1C) over consecutive 24-hour periods compared to adult rats. The lower water intake in aged rats was not accompanied by a significant decrease in urine output compared to adult rats (Fig. 1D) but reflected a decrease in extrarenal water loss compared to their younger counterparts (Fig. 1E). Urine osmolality was not affected by age (Fig. 1F).

3.2. The AVP system in adult and aged rats

To test how fluid homeostatic systems respond to osmotic stress, rats were deprived of water for 3 days. A decrease in weight from the starting body weight was observed for both adult and aged rats (Fig. 2A). This period of dehydration increased plasma osmolality by similar degrees in both aged and adult rats (Fig. 2B). However, a higher basal plasma osmolality was observed in aged rats, a difference that was preserved in response to 3 days of dehydration, suggestive of different osmolality set points in adult and aged rats. We investigated the expression of Avp mRNA and hnAvp RNA, a surrogate measure of Avp transcription (Herman et al., 1991), in the SON of control and dehydrated adult and aged rats using qRT-PCR (Fig. 2C and D). The abundance of Avp mRNA under basal conditions was not influenced by age (Fig. 2C), while increased hnAvp expression in aged animals indicated increased transcription of the Avp gene compared to adult rats (Fig. 2D). In contrast, the osmotic stimulus of dehydration increased hnAvp levels above adult basal measures for both adult and aged rats; however, this response was only significant in adult rats. The AVP content in the pituitary was investigated by AVP ELISA (Fig. 2E). There was a decrease in NIL AVP content in rats subjected to dehydration for both age groups. AVP NIL content was unchanged by age. Interestingly, the expected decline of AVP content with dehydration was marginally attenuated in aged rats,

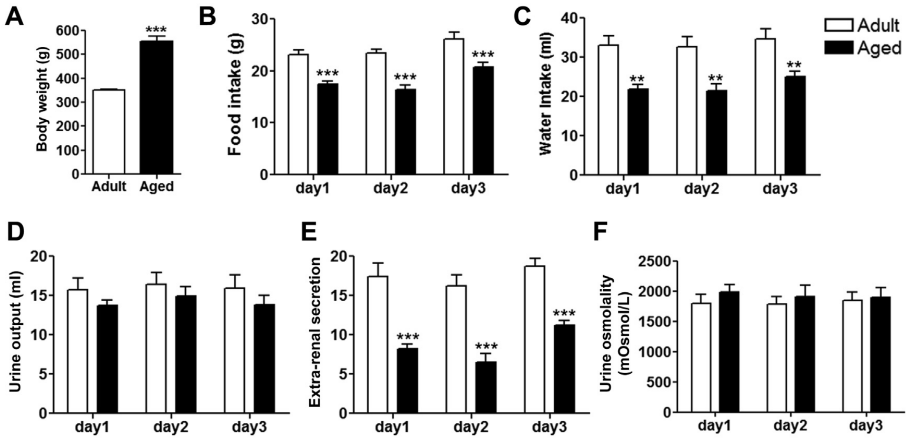


Fig. 1. Comparison of metabolic parameters in adult and aged rats. Physiological parameters—(A) body weight, (B) food intake, (C) water intake, (D) urine output, (E) extrarenal water secretion, and (F) urine osmolality—were recorded in adult (3-month old) and aged (18-month old) male rats (n = 6) housed in metabolic cages. **, $p < 0.01$; ***, $p < 0.001$ by the 2-way analysis of variance with Bonferroni post hoc test.

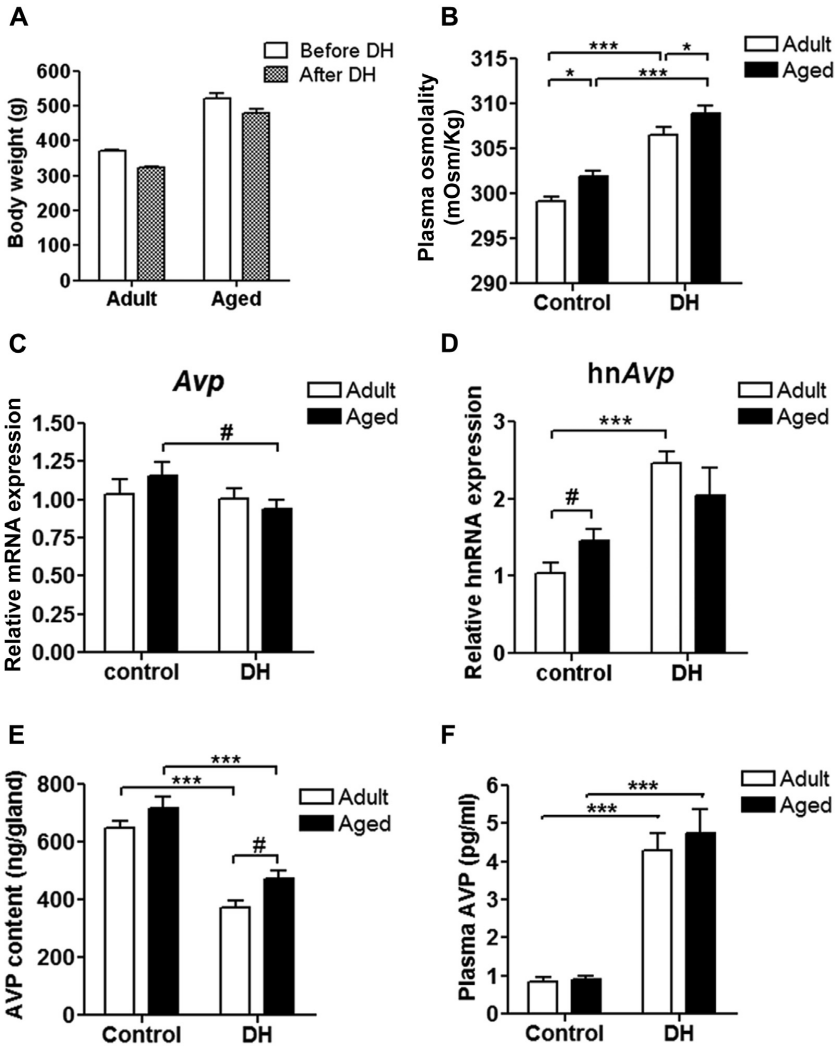


Fig. 2. Effect of aging on the rat AVP system. Adult and aged male rats were subjected to dehydration for 3 days and AVP measures were performed. (A) Body weights were recorded before and after dehydration (n = 11–14). (B) Plasma osmolality was measured (n = 19–23) by freezing point depression. The effect of aging on Avp expression was examined at the transcriptional level in control and dehydrated adult and aged rats. (C, D) The RNA expression level of Avp (both heteronuclear [hnAvp] and mature form) was examined by qRT-PCR (n = 6). AVP measures were performed on NILs and plasma extracts. (E) AVP content in NILs was measured by ELISA (n = 9–10). (F) Plasma AVP level was determined by radioimmunoassay (n = 9). *, $p < 0.05$; ***, $p < 0.001$ by the 2-way analysis of variance with Bonferroni post hoc test. #, $p < 0.05$ by unpaired *t*-test. Abbreviations: AVP, arginine vasopressin; DH, dehydrated; NIL, neurointermediate lobe.

with higher AVP levels detected in aged dehydrated compared to adult dehydrated rats. However, there was no significant effect of age on basal or osmotically induced plasma AVP levels when comparing adult and aged rats (Fig. 2F).

3.3. Peptide analysis of the NIL in control and dehydrated states

3.3.1. Effect of dehydration on NIL peptide profiles in adult rats

The abundance of peptides in the NIL provides a good measure for assessing changes in peptide synthesis/secretion. To study effects of aging on peptide profiles in the NIL, we used MS-based peptide measurements (Romanova et al., 2013; Romanova and Sweedler, 2015) to characterize and quantify the neuropeptide changes of the NIL. In adult rats, control and dehydrated profiles were easily classified by PCA according to principal component 1 (PC1) (~60% of variance; Fig. 3A). Loading plots indicated that peptides contributing to this difference matched the masses of AVP and its sodiated ion, the sodiated ion of oxytocin (OT), acetylated alpha-MSH, di-acetylated alpha-MSH, a portion of the ACTH domain, and other proopiomelanocortin (POMC)-derived peptides (Fig. 3B). The level of AVP and OT decreased with dehydration, whereas alpha-MSH and proSAAS levels increased (Fig. 3C).

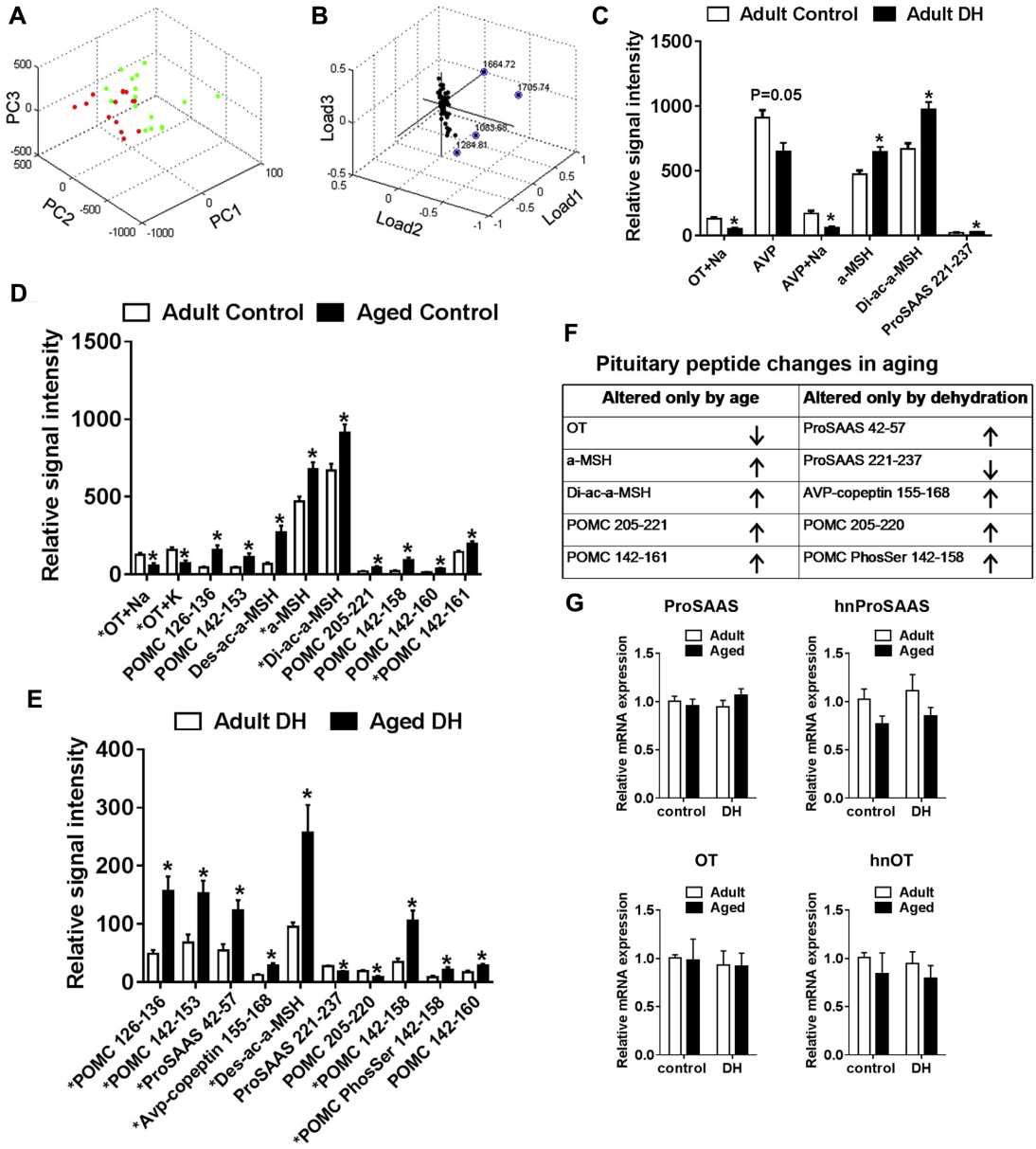


Fig. 3. Peptide analysis in aged rats. Adult and aged male rats were subjected to dehydration for 3 days, and NIL peptide measures were performed by MALDI-TOF MS. (A–E) Signal intensity of peptides in NILs was measured by MALDI-TOF MS in individual tissue extracts from adult and aged rats in basal and dehydrated states (n = 11–16). (A) Control and dehydrated peptide profiles are easily classified by PCA according to PC1 (~60% of variance) in adult rats; PCA plot is shown for the first 3 PCs. (B) Loading plot indicates that a small subset of peptides contributes to differences between control and dehydrated adult rats, among which are masses matching AVP, acetylated alpha-MSH, and di-acetylated alpha-MSH. (C) The AVP and OT signal decreases significantly, while the alpha-MSH signals and proSAAS (221-237) increase significantly with dehydration in adult rats. (D, E) The effect of age on peptide signals in the NIL. (D) The OT signals decreased, while the signals of POMC-derived peptides increased in aged compared to adult rats. (E) Peptide signals in the NIL of adult and aged rats subjected to 3 days of dehydration. The AVP-copeptin signal increased, whereas the proSAAS signal decreased with age. (F) Peptides uniquely altered as a function or aging or by just dehydration. (G) Relative RNA expression of proSAAS and OT in the SON of adult and aged rats. *$p < 0.05$ by unpaired *t*-test or peptide names beginning with * by Kruskal-Wallis tests. Abbreviations: AVP, arginine vasopressin; DH, dehydrated; NIL, neurointermediate lobe; PCA, principal component analysis; POMC, proopiomelanocortin.

3.3.2. Comparison of the NIL peptide profiles between aged and adult rats

A comparison was performed for adult and aged rats. In PCA, 10 PCs were required to explain 93% of variance in the data set, with most sample segregation achieved along PC1 accounting for about 30% of variance. Spectra from aged animals showed more broad distribution within the 3D space constructed of the first 3 PCs. In this data set, a total of 78 peaks passed the criteria for statistics selection, of which 22 (~30%) were detected at statistically different intensities ($p \leq 0.05$) between compared age groups (see Supplementary Tables 1 and 2). Some of the peaks can be matched to POMC by peptide mass fingerprinting or other previously reported neuropeptides expressed in pituitary including OT (Fig. 3D). Relative to adult rats, the aged rats exhibited significant decrease in the sodiated ion and potassiumated ion of OT and increases in the intensity of peptides matching the masses of alpha-MSH, acetylated alpha-MSH, di-acetylated alpha-MSH, and 5 other POMC-derived peptides.

3.3.3. Effect of aging on NIL peptide profiles in dehydrated rats

A comparison was performed for NILs of adult and aged dehydrated rats. In PCA, 13 PCs were required to explain 95% of variance in the data set. Both adult and aged rats showed a range of profiles that could not be reliably classified by PCA and unsupervised clustering. Four of 13 adult rats and 2 out of 16 aged rats were particularly different. A total of 121 peaks were selected for statistics, of which 20 (~16%) were detected at statistically different intensities ($p \leq 0.05$) between compared age groups (see Supplementary Tables 3 and 4). Similar to the control groups, aged animals had higher levels of peptides matching by mass to the POMC prohormone as well as proSAAS and AVP-copeptin (Fig. 3E).

3.3.4. Effect of dehydration on NIL peptide profiles in aged rats

With the set of 24 aged rats (11 control and 13 dehydrated), no significant changes in NIL profiles were seen with dehydration, and no clear segregation was observed on a PCA plot.

3.3.5. Peptide changes specific to age or dehydration

A number of peptides profiles in the NIL were altered only as a function of aging (Fig. 3F). In addition, a separate cohort of peptides was found to only differ between adult and aged rats in dehydration. (Fig. 3F). These included OT (aging) and proSAAS (dehydration) whose precursor proteins are known to be synthesized in magnocellular neurons in the SON (Murphy et al., 2012). Using qRT-PCR, we show that expression of these genes in the SON is not altered by age or dehydration (Fig. 3G).

3.4. Changes in Avp promoter methylation as a consequence of aging

To see if changes in methylation could account for Avp gene transcriptional differences in the SON with age, we looked at the expression of genes known to regulate methylation status of DNA, namely the Dnmt and Tet families, in the SON (Fig. 4A). We found decreased Dnmt1 and Tet1 in the SON of aged compared to adult rats, whereas expression of the closely related genes Dnmt3a and Tet2/3 remained unchanged with age. In the dehydrated state, Dnmt1 increased and Tet1 decreased in adult rat SON samples, whereas no changes in these genes were observed with dehydration and aging. To analyze gene-specific methylation changes, we chose to examine the methylation profile of the Avp promoter within the SON by sequence analysis of bisulfite-converted DNA. Using primers spanning the proximal Avp promoter (−325 to −24 bp), we investigated the methylation status of a cluster of 7 cytosine-phosphate-guanine (CpG) sites (Fig. 4B). Analyses of the methylation pattern of CpGs in single clones from individual control and dehydrated animals with aging are depicted in Fig. 4C. Analysis of the overall methylation of the Avp promoter for the SON revealed decreased methylation in aged compared to adult animals by the 2-way analysis of variance ($p < 0.002$), whereas methylation levels increased in response to dehydration in aged rats (Fig. 4D). In comparison, overall methylation was not significantly altered by dehydration in the SON of adults.

We next compared the methylation profiles of individual CpGs (Fig. 4E). Of the 7 CpGs analyzed, only CpG2 was significantly influenced by age, with lower methylation compared to adult controls. In aged rats, dehydration increased methylation of CpGs 1, 3–5, and 7 compared to aged controls. By contrast, only CpG4 showed increased methylation in dehydrated adult rats compared to adult controls. Of note, the methylation of individual CpGs was found to be similar in dehydrated adult and aged rats.

3.5. Aging changes gene expression in the SON and alters the effect of dehydration

We have used transcriptomics to catalog all of the genes expressed in the adult male SON and to identify genes that are differentially regulated by dehydration (Hindmarch et al., 2006). The challenge now is to place these genes into physiologically relevant pathways; thus, in pursuit of this aim, our functional investigations have revealed novel genes involved in AVP elaboration [Creb3l1 (Greenwood et al., 2014, 2015a,b); Slc12a1 (Konopacka et al., 2015b); Caprin2 (Konopacka et al., 2015a); Giot1 (Qiu et al., 2007); and Rasd1 (Greenwood et al., 2016b)]. We have now asked if the expression of these genes is altered with aging, under both euhydrated and dehydrated conditions (Fig. 5). We used qRT-PCR to reveal age-related increases in mRNA expression of transcription factors c-Fos (a general marker of neuronal activation), Creb3l1, Giot1, and RNA-binding protein Caprin2 under basal conditions, while levels of the small G-protein Rasd1 and the Na-K-2Cl cotransporter Slc12a1 were unchanged. The expression of all of these genes was increased by dehydration in both adult and aged animals. In aged rats, dehydration induced smaller rises in the expression of all analyzed genes, reaching statistical significance compared to adult dehydrated rats, with 1 notable exception, Caprin2.

3.6. Post-transcriptional modification to Avp mRNA in aging

A known feature of the Avp mRNA is that the unusually long 3′ poly(A) tail further increases in length in response to osmotic stress (Carter and Murphy, 1991). Here, we have used poly(A) tail assays to determine the length of the Avp mRNA poly(A) tail in adult and aged rats in the basal condition and in response to dehydration (Fig. 6A). The length of the Avp poly(A) tail was found to be susceptible to change with aging. In aged rats, the poly(A) tail was longer than that in adult control rats (Fig. 6B), perhaps suggesting altered transcript stability with age. The Avp poly(A) tail length increased more in adult rats in dehydration, but overall poly(A) tail lengths ended up being the same size in both dehydrated groups reflecting the smaller starting point in adult rats.

4. Discussion

With increased life expectancy, maintaining health and well-being into old age is becoming a priority, making the push toward understanding our aging homeostatic systems ever more pertinent. A decline of appetite accompanied by a reduction in daily fluid intake, as we observed in the rat, is common behavioral characteristics observed in the elderly (Kmiec, 2006; Phillips et al., 1993),

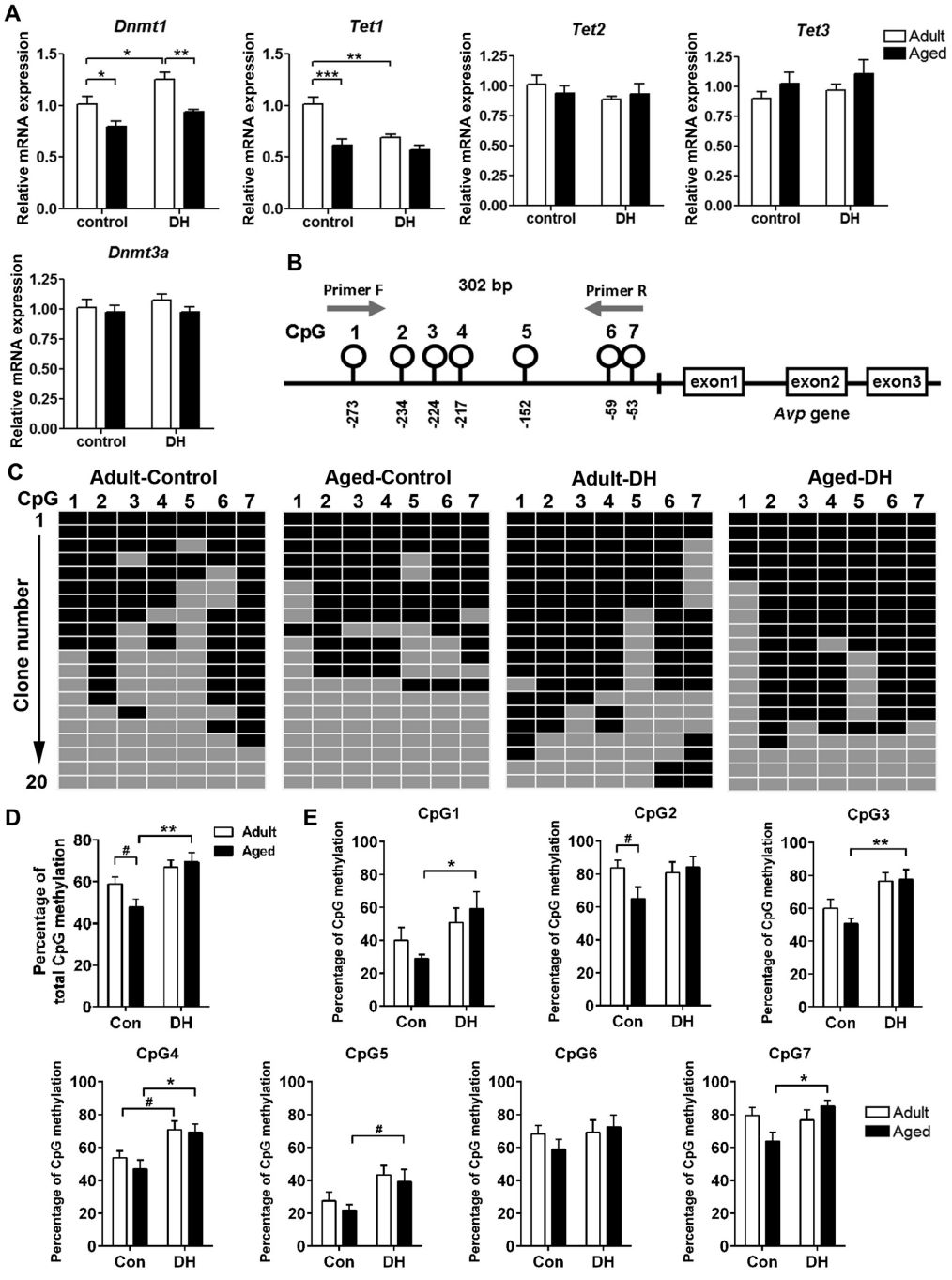


Fig. 4. Epigenetic changes in Avp gene promoter in aging. Methylation status of the Avp promoter in the SON of control and dehydrated adult and aged male rats. (A) Relative mRNA expression of Dnmt1, Dnmt3a, Tet1, Tet2, and Tet3 was determined by qRT-PCR. (B) Diagram showing 7 CpG sites on the Avp promoter that were examined by colony-based PCR. (C) Representative tile diagrams showing the methylation status of 7 CpG sites for individual clones of the Avp promoter extracted from the SON. (D) Percentage of global methylation on this region of the Avp promoter in control and dehydrated adult and aged rats. (E) Percentage methylation of individual CpG sites on the Avp promoter in control and dehydrated adult and aged rats. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$ by the 2-way analysis of variance with Tukey's post hoc test. #, $p < 0.05$ by unpaired *t*-test. Abbreviation: SON, supraoptic nucleus.

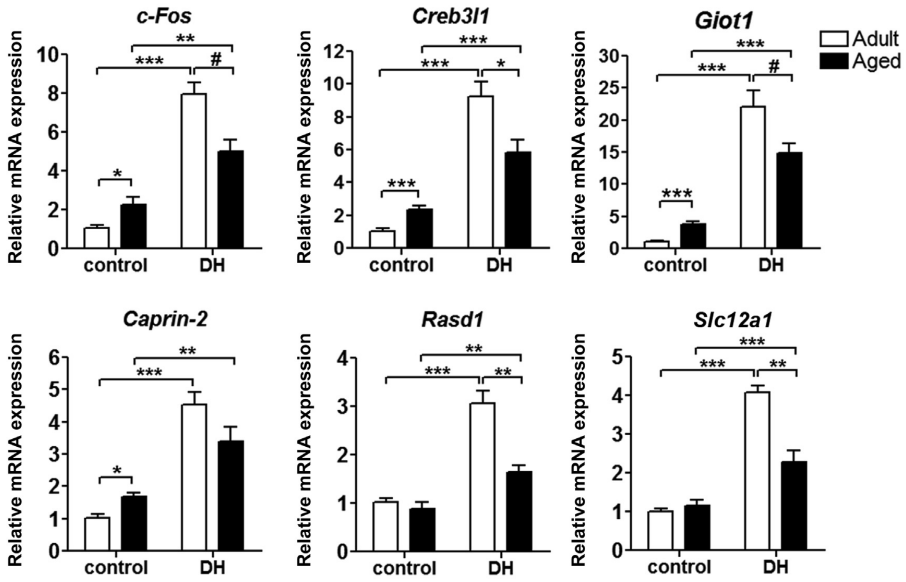


Fig. 5. Effect of aging on the expression of osmotically induced genes. Relative mRNA expression of genes involved in hyperosmotic stress in the SON of the hypothalamus in control and 3-day dehydrated adult and aged rats. *, $p < 0.05$; **, $p < 0.01$; and ***, $p < 0.001$ by the 2-way analysis of variance with Bonferroni post hoc test. #, $p < 0.05$ by unpaired *t*-test. Abbreviations: DH, dehydrated; SON, supraoptic nucleus.

suggesting that our rat model is of particular value to study metabolic changes related to aging. Comparisons of basal and dehydrated urine osmolalities in both age groups suggested that urine concentrating capacity and thus renal function are not impaired in our model at this age. Furthermore, AVP circulating levels were comparable in adult and aged rats in the basal state and in response to dehydration, suggesting no changes in hypothalamo-neurohypophysial system (HNS) responsiveness to osmotic and volume stimuli. Hence, the circulating levels of AVP and the renal response to it are adequate to concentrate urine in these aging rats.

Chronic dehydration depletes AVP stores in the posterior pituitary to meet necessary circulatory demands for AVP to facilitate increased water uptake by the kidney (Antunes-Rodrigues et al., 2014). The large stores of AVP in the posterior pituitary were depleted by comparable amounts in both age groups by dehydration. However, when compared to adult rats, aged rats had higher levels of pituitary AVP as a consequence of dehydration. We propose that increased basal AVP pituitary is responsible for this difference as opposed to ineffective stimulation of AVP secretion. This concept is consistent with the comparable increases in circulating levels of AVP after dehydration. Therefore, the ability to store and secrete adequate quantities of AVP in response to 3 days of dehydration was not compromised in these Han Wistar rats at this age. Taken together, these data show that changes to AVP secretion cannot account for the observed metabolic changes in aged compared to adult rats. The altered fluid intake and plasma electrolytes may represent changes to other systems coordinating salt and water balance. For example, the renin-angiotensin-aldosterone and atrial natriuretic peptide systems are known to be altered in rats and humans as a function of aging (El-Sharkawy et al., 2014; Pollack et al., 1997; Silver et al., 1993). However, any involvement of these systems in this particular aging model remains to be investigated.

A higher set-point for basal plasma osmolality in our aged model, one of the reported characteristics of aging in humans and rodents (McLean et al., 1992; Terwel et al., 1992), provides one possible explanation for AVP neuron hyperactivation in the basal state. A small rise in plasma osmolality of approximately 1% is normally sufficient to activate Avp transcription in magnocellular neurons of the SON and PVN, and these transcriptional events are well known to occur together with increased AVP secretion from the posterior pituitary in adult rats (Arima et al., 1999). The higher plasma osmolality in aged rats, being approximately 1% above adult rats, was indeed associated with increased transcription but not with increased secretion of AVP. In contrast, 3 days of dehydration, a well-characterized model for activating Avp transcription in the SON (Greenwood et al., 2014), increased AVP secretion but not Avp transcription in the aged group. This is despite a rise in plasma osmolality of greater than 2% by this osmotic stimulus.

The synthesis and secretion of AVP are normally twinned to maintain neurohypophysial homeostasis because AVP stores in the pituitary become depleted and need to be replenished with newly synthesized AVP (Murphy and Carter, 1990). Any delay in replenishing pituitary AVP stores to prestimulus levels might leave the system at greater risk from further hyperosmotic insults. The elderly living at care homes have been shown to have lower daily intakes of fluid than those living at home. Furthermore, elderly people with cognitive impairments such as dementia often forget to drink. These behavioral characteristics, coupled with reduced thirst perception in elderly people, greatly increase their risk of dehydration (El-Sharkawy et al., 2014). In elderly patients admitted to hospitals, hypernatremia has been associated with an increased mortality rate (Snyder et al., 1987). In addition, clinical studies of care home patients who develop acute illness and require hospital treatment reported that approximately 30% became markedly hypernatremic in hospitals (Cowen et al., 2013). The uncoupling of plasma osmolality and Avp transcription did not alter AVP secretion here in healthy aging rats but may become important in

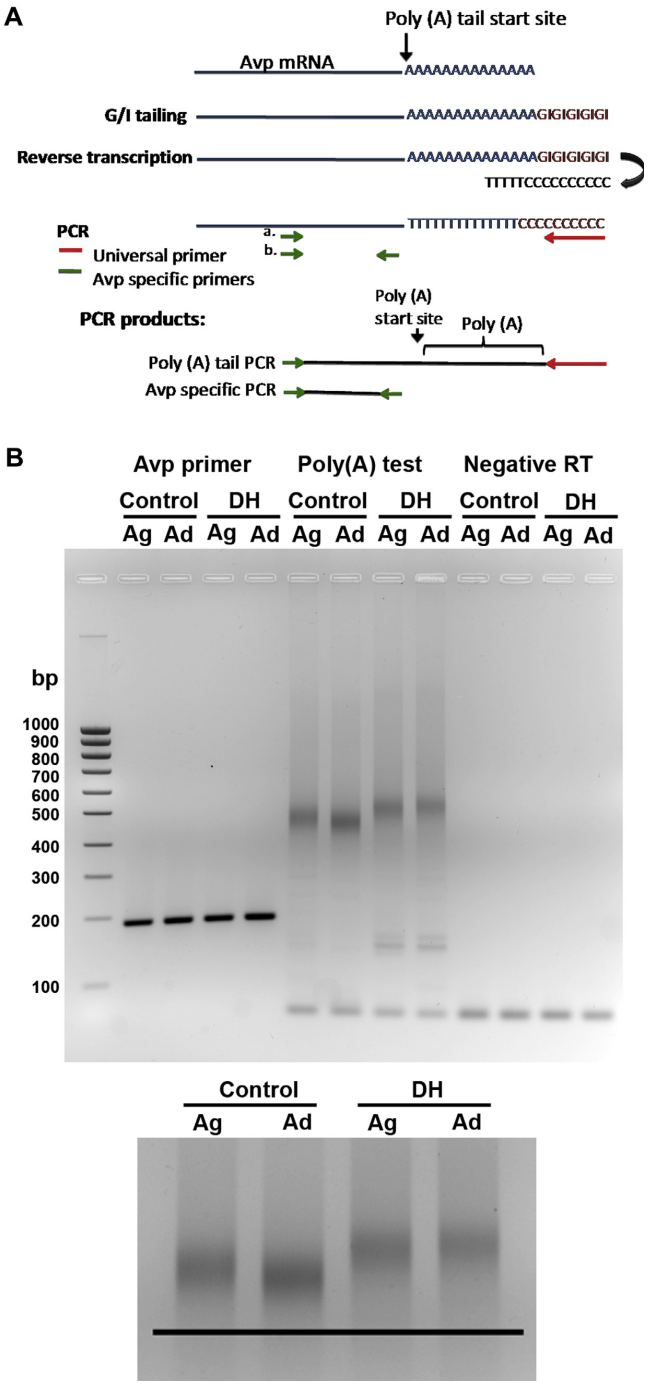


Fig. 6. Effect of aging on transcriptional and post-transcriptional Avp gene expression. The effect of aging on Avp poly(A) tail length in the SON in control and dehydrated adult and aged rats. (A) Diagram of poly(A) tail assay design for the rat Avp gene. (B) poly(A) tail length of the Avp mRNA was examined using a PCR-based poly(A) tail assay. Abbreviations: Ad, adult; Ag, aged; DH, dehydrated; RT, reverse transcription; SON, supraoptic nucleus.

pathophysiological conditions if the rate of Avp transcription ever fails to meet secretory demands.

We next investigated if the observed uncoupling of synthesis and secretion was unique to AVP in the aging HNS. We revealed a cohort of peptides, in particular, POMC-derived peptides, N-terminal truncated form of copeptin, proSAAS, and OT in addition to AVP, which were susceptible to changes with age and also dehydration. POMC has been shown to be expressed in the pituitary intermediate lobe, with its expression altered by osmotic stimulation and changes in blood pressure (Felder and Garland, 1989; Pardy et al., 1990). In rats supplied with a drinking diet of 2% NaCl, POMC mRNA expression was shown to decrease in the intermediate lobe of the pituitary (Pardy et al., 1990). In relation to blood pressure, spontaneously hypertensive (SHR) rats have lower POMC expression in the intermediate lobe compared to Wistar Kyoto (WKY) rats. Lowering blood pressure in SHR rats with antihypertensive agents normalizes POMC expression in the intermediate lobe to that in the WKY rat (Felder and Garland, 1989). We show here age-related increases in an array of POMC-derived peptides in the NIL under basal and dehydrated conditions. How, or if, these POMC peptides contribute to age-related changes to physiology is currently not known.

Copeptin and AVP are derived from the same common precursor molecule. Copeptin is the C-terminal part of pro-AVP that is cleaved during processing and released with AVP into the circulation. The functions of copeptin are not known, but due to its higher stability in plasma, it is commonly used as surrogate measure for circulating levels of AVP (Christ-Crain and Fenske, 2016). Here, we have identified an N-terminal truncated form of copeptin in the NIL and further show that the abundance of this peptide increases in the dehydrated rat as a function of aging. This peptide has previously been identified in a peptidomic study of the rat SON (Bora et al., 2008). Therefore, increased abundance of this peptide in the NIL might suggest increased processing of pro-AVP in aging magnocellular neurons in response to dehydration. In support of this concept, a study using microdialysis probes to measure release patterns of AVP in the PVN and SON in the aging male Wistar rat showed an age-associated increase in AVP release in the PVN, though not in the SON (Keck et al., 2000).

ProSAAS and OT are known to be expressed in magnocellular neurons of the hypothalamus and the posterior lobe of the pituitary gland (Bora et al., 2008; Gouraud et al., 2007). The propeptide precursor ProSAAS is processed into a number of smaller peptides in the brain and pituitary including big SAAS, little SAAS, PEN, big LEN, and little LEN (Mzhavia et al., 2001). Here, we identify age-associated alterations in truncated forms of little SAAS (ProSAAS 42-57) and PEN (ProSAAS 221-237) in the dehydrated NIL. Interestingly, dehydration for 3 days also increases ProSAAS expression in the SON while decreasing ProSAAS expression in the NIL, suggesting that the ProSAAS peptide might be secreted (Gouraud et al., 2007), although this remains to be determined. The previous peptidomic study performed on rat SON samples also identified this cleavage of ProSAAS (42-57) and multiple PEN peptides in magnocellular neurons. One possibility for the involvement of ProSAAS in the regulation AVP is through its interactions with proprotein convertase 1. Pro-AVP is processed by proprotein convertase 1, and ProSAAS inhibits the activity of this convertase (Murphy et al., 2012). Therefore, changes in ProSAAS expression in aging might alter the processing and thus the availability of AVP.

OT is probably best known for its role in female reproduction but also functions as a natriuretic hormone that reduces sodium appetite and increases sodium excretion at the kidney (Verbalis et al., 1991). Circulating levels of OT increase in response to stimulation by osmotic stress, and this depletes pituitary stores of this peptide consistent with this study (Silverman et al., 1990). Our data suggest that OT stores are depleted in the NIL as a consequence of aging, in agreement with earlier studies of the aging HNS (Silverman et al., 1990; Zbuzek et al., 1988). Because OT and ProSAAS are synthesized in magnocellular neurons of the hypothalamus, such as AVP, we have also investigated their transcript abundance in the SON to assess this phenomena of uncoupling observed for AVP. We report no age-related effects on OT or ProSAAS transcription in the SON, although our data indicate altered peptide levels in the pituitary.

We have made significant steps to understanding the mechanism regulating AVP in the adult rat and have applied our understanding to the Avp transcriptional changes in this aging model. Based on our recent study where we described altered methylation patterns of the Avp promoter with dehydration, we hypothesized that altered methylation marks in the Avp promoter could be responsible. Cellular aging is closely associated with a decrease in expression of Dnmt1, an enzyme that stabilizes methylation marks on DNA (Casillas et al., 2003), as we observed here in the SON. This is thought to be 1 reason for hypomethylation of DNA sequences in rodents as well as humans and is consistent with hypomethylation of the Avp promoter in the aged rat SON. We also found lower levels of Tet1 expression in the aged rat and adult SON following dehydration. Tet1, by hydroxylation of 5hmc, has been shown to promote active demethylation of DNA in the rodent brain (Guo et al., 2011).

The decrease in Avp promoter methylation in aged rats may perhaps explain the increased Avp transcription in the aging SON. Many studies have shown that lower levels of promoter methylation correlate with increased gene transcription. We previously reported increased Avp transcription in hypothalamic 4B cells following demethylation by 5-aza-2′-deoxycytidine treatment, consistent with this hypothesis (Greenwood et al., 2016a). Individual CpG sites were largely unaffected by age, apart from CpG2, which resides close to a cAMP responsive element, which underwent hypomethylation with age. Methylation at cAMP response element sites has been shown to inhibit cAMP response element binding protein (CREB)–mediated transcription (Elliott et al., 2010; Zhang et al., 2005), so hypomethylation of this site may serve to enhance Avp promoter activation by CREB (Iwasaki et al., 1997). Interestingly, methylation signatures on this segment of the Avp promoter remained largely unchanged by dehydration in adult rats, differing from our findings in the Sprague Dawley rat (Greenwood et al., 2016a). We suggest this is due to strain differences. Nonetheless, dehydration induced the hypermethylation of CpG3, 4, and 7 in aged rats, as observed in adult Sprague Dawley rats, thus adding strength to the argument for a relationship between the methylation status of specific Avp promoter CpGs and Avp gene transcription.

To further aid our understanding of AVP neuron activity in aging, we looked at the expression of genes known to be robustly induced by osmotic stimuli, and whose functions have been the subject of interrogation by us in relation to AVP biosynthesis in the SON, and the overall regulation of fluid balance in the rat (Greenwood et al., 2014, 2016b; Konopacka et al., 2015a,b; Qiu et al., 2007). We recently identified Creb3l1 as a putative transcription factor regulating the expression of the Avp gene (Greenwood et al., 2014). Therefore, in the basal condition, increased Avp transcription can perhaps be explained by the upregulation of Creb3l1 expression, and conversely, the attenuated Creb3l1 induction in aged dehydrated rats following osmotic stimulation may explain the reduced capacity to elevate Avp. The altered expression of genes regulating transcriptional events (c-Fos, Creb3l1, and Giot1) in basal and dehydrated states implies dramatic changes in the SON transcriptome with aging. These genes are all activated by cAMP pathways (Greenwood et al., 2015a; Qiu et al., 2007), suggesting that altered cAMP signaling may determine altered transcriptional

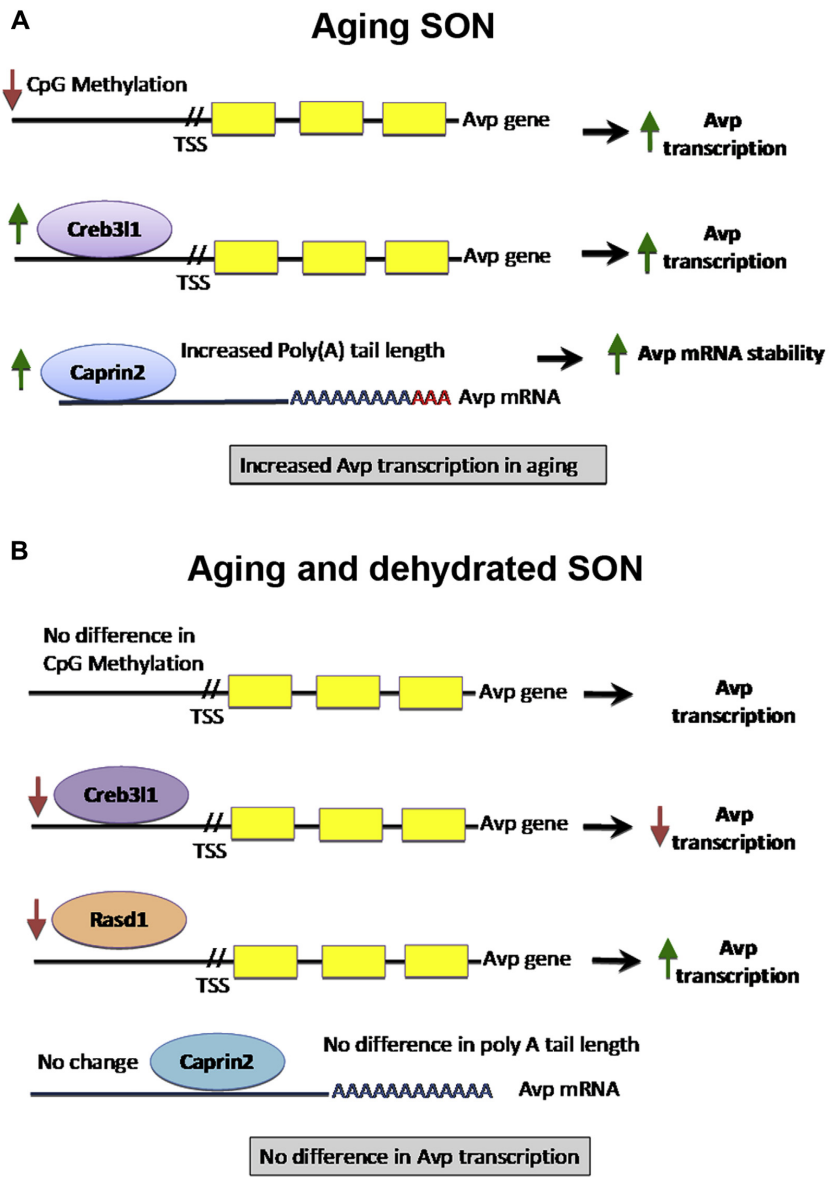


Fig. 7. Modeling of molecular events in the SON in aging and dehydrated rats. (A) How aging may alter Avp expression in the SON compared to younger counterparts. (B) Proposed differential regulation of Avp expression in the SON as a function of both aging and dehydration. Abbreviation: SON, supraoptic nucleus.

responses in the aged SON. The source of these altered signaling responses is not known but may be a consequence of either altered inputs from the circumventricular organs due to changes in plasma osmolality (McKinley et al., 2004) or changes within the magnocellular neurons themselves.

It is interesting to note that basal Rasd1 and Slc12a1 expression levels were not influenced by age implying activation of these genes by separate signaling pathways not affected by age. Rasd1 is a member of the Ras family of small G-proteins, which is expressed in AVP magnocellular neurons of the PVN and SON, where increased circulating glucocorticoids and/or raised plasma osmolality induce its expression (Greenwood et al., 2016b). Interestingly, plasma corticosterone levels increase in aging rodents as a result of hyperactivity of the hypothalamo-pituitary-adrenal axis (Goncharova, 2013). At the same time, there is an age-associated decrease in the sensitivity of the hypothalamus, along with other

brain nuclei, to glucocorticoids (Goncharova, 2013), which may account for the blunted increase in Rasd1 expression in dehydrated aged animals, despite changes to corticosterone levels. By lentiviral vector–mediated overexpression of Rasd1 in the SON, we recently showed that Rasd1 inhibits osmotically induced Avp transcription in this nucleus (Greenwood et al., 2016b). The aged dehydrated rats appear to have lost this dehydration-induced inhibitory input on Avp transcription in the aging SON.

The expression of Slc12a1 in magnocellular neurons of the SON and PVN is also known to be induced by chronic and acute osmotic stimulation (Konopacka et al., 2015b). We recently showed that lentiviral-mediated knockdown of Slc12a1 in these hypothalamic nuclei altered fluid homeostasis by increasing fluid intake and urine output during salt loading. Furthermore, the loop diuretics bumetanide and furosemide were found to inhibit gamma-aminobutyric acid-mediated excitation of AVP neurons and AVP release, respectively. Therefore, an altered abundance of Slc12a1 might be expected to alter neuronal activity in aging AVP neurons.

The Avp mRNA is subject to post-transcriptional modification in the form of an increase in the length of the poly(A) tail, as seen in the SON in response to osmotic challenges (Carter and Murphy, 1991) and as we show here by age. An increased poly(A) tail length is thought to reduce the degradation and increase the stability of many transcripts (Zeevi et al., 1982) and may be involved in the control of translation (Palatnik et al., 1984). We recently showed that the RNA-binding protein Caprin2 binds to the Avp mRNA and, in doing so, mediates an increase in the length of the poly(A) tail (Konopacka et al., 2015a). It is interesting to note that the expression of Caprin2 in the aging SON increases in parallel with the Avp mRNA poly(A) tail length. The increase in Avp mRNA poly(A) tail length in response to dehydration was not affected by age, as previously reported (Sladek and Olschowka, 1994), which was further corroborated by there being no difference in Caprin2 expression in dehydrated adult and aged rats. A previous Northern blot study on Fischer 344 rat SON samples reported no change in Avp poly(A) tail length with age (Sladek and Olschowka, 1994). We suggest that this discrepancy could be due either to strain differences or to the different methodological approaches used.

4.1. Conclusions

In summary, we have performed a comprehensive analysis of the AVP system in aging Han Wistar rats. There were no age-related changes to AVP circulating levels in basal or dehydrated states, suggesting that the functioning of the HNS in body water homeostasis is intact in healthy rats at this age. In stark contrast, we describe in the magnocellular neurons of the SON a plethora of molecular changes known to alter Avp expression. These include methylation changes to the Avp promoter and altered expression of genes involved in transcriptional and post-transcriptional regulation of the Avp gene (Fig. 7). We currently do not understand the origin of these changes or why they are a necessary part of normal aging in the rat. The current rat model was perfectly capable of coping with 3 days of dehydration by secreting adequate quantities of AVP. However, this may not be true for pathophysiological conditions commonly encountered in the aging process. The stimulus secretion uncoupling of Avp transcription as seen here in normal aging could increase the likelihood of fluid and electrolyte disorders in critically ill elderly patients. Not being able to adequately correct or respond to fluid and electrolyte disturbances such as a rise plasma osmolality would certainly lead to hypernatremia, a condition that is regularly observed in elderly patients admitted to hospitals (El-Sharkawy et al., 2014). Further studies are necessary to address whether Avp transcription can be a rate-limiting step in the aging HNS. The identification by MS of additional peptides that are also influenced by age in the NIL gives further scope for exploring other peptidergic systems in the aging hypothalamus, in the context of healthy aging. Understanding how aging alters these hormonal systems may help to improve treatment regimens and perhaps improve the clinical outcomes for elderly patients.

Disclosure statement

The authors have no actual or potential conflicts of interest.

Acknowledgements

The authors gratefully acknowledge the support of the MRC (MR/N022807/1 to MG and DM), the BBSRC (BB/G006156/1 to MPG, OS, and DM; BB/J015415/1 to MG and DM; BB/J01981X/1 to DM and JVS), a High Impact Research Chancellory Grant from the University of Malaya (UM.C/625/1/HIR/MOHE/MED/22 H-20001-E000086 to ASM, OS, and DM), and Sao Paulo Research Foundation FAPESP (2013/09799-1 to JAR and ASM). This work was also supported by Award Number P30 DA018310 (JVS and EVR) from the National Institute on Drug Abuse (JVS). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.neurobiolaging.2018.01.008.

References

Antunes-Rodrigues, J., Ruginsk, S.G., Mecawi, A.S., Margatho, L.O., Reis, W.L., Ventura, R.R., da Silva, A.L., Vilhena-Franco, T., Elias, L.L.K., 2014. Neuroendocrinology of hydromineral homeostasis. In: De Luca Jr., L.A., Menani, J.V., Johnson, A.K. (Eds.), Neurobiology of Body Fluid Homeostasis: Transduction and Integration. CRC Press/Taylor and Francis, Boca Raton (FL).

Ares, G.R., Caceres, P.S., Ortiz, P.A., 2011. Molecular regulation of NKCC2 in the thick ascending limb. Am. J. Physiol. Ren. Physiol. 301, F1143–F1159.

Arima, H., Kondo, K., Kakiya, S., Nagasaki, H., Yokoi, H., Yambe, Y., Murase, T., Iwasaki, Y., Oiso, Y., 1999. Rapid and sensitive vasopressin heteronuclear RNA responses to changes in plasma osmolality. J. Neuroendocrinol. 11, 337–341.

Bora, A., Annangudi, S.P., Millet, L.J., Rubakhin, S.S., Forbes, A.J., Kelleher, N.L., Gillette, M.U., Sweedler, J.V., 2008. Neuropeptidomics of the supraoptic rat nucleus. J. Proteome Res. 7, 4992–5003.

Breyer, M.D., Ando, Y., 1994. Hormonal signaling and regulation of salt and water transport in the collecting duct. Annu. Rev. Physiol. 56, 711–739.

Burke, S.N., Barnes, C.A., 2006. Neural plasticity in the ageing brain. Nat. Rev. Neurosci. 7, 30–40.

Carter, D.A., Murphy, D., 1991. Rapid changes in poly (A) tail length of vasopressin and oxytocin mRNAs form a common early component of neurohypophyseal peptide gene activation following physiological stimulation. Neuroendocrinology 53, 1–6.

Casillas Jr., M.A., Lopatina, N., Andrews, L.G., Tollefsbol, T.O., 2003. Transcriptional control of the DNA methyltransferases is altered in aging and neoplastically-transformed human fibroblasts. Mol. Cell Biochem. 252, 33–43.

Christ-Crain, M., Fenske, W., 2016. Copeptin in the diagnosis of vasopressin-dependent disorders of fluid homeostasis. Nat. Rev. Endocrinol. 12, 168–176.

Cowen, L.E., Hodak, S.P., Verbalis, J.G., 2013. Age-associated abnormalities of water homeostasis. Endocrinol. Metab. Clin. North Am. 42, 349–370.

Dudoit, S., Shaffer, J.P., 2003. Multiple hypothesis testing in microarray experiments. Stat. Sci. 18, 71–103.

El-Sharkawy, A.M., Sahota, O., Maughan, R.J., Lobo, D.N., 2014. The pathophysiology of fluid and electrolyte balance in the older adult surgical patient. Clin. Nutr. 33, 6–13.

Elliott, E., Ezra-Nevo, G., Regev, L., Neufeld-Cohen, A., Chen, A., 2010. Resilience to social stress coincides with functional DNA methylation of the Crf gene in adult mice. Nat. Neurosci. 13, 1351–1353.

Felder, R.A., Garland, D.S., 1989. POMC biosynthesis in the intermediate lobe of the spontaneously hypertensive rat. Am. J. Hypertens. 2, 618–624.

Frolkis, V.V., Kvitnitskaya-Ryzhova, T.Y., Dubiley, T.A., 1999. Vasopressin, hypothalamo-neurohypophyseal system and aging. Arch. Gerontol. Geriatr. 29, 193–214.

Goncharova, N.D., 2013. Stress responsiveness of the hypothalamic-pituitary-adrenal axis: age-related features of the vasopressinergic regulation. Front. Endocrinol. (Lausanne) 4, 26.
Gouraud, S.S., Heesom, K., Yao, S.T., Qiu, J., Paton, J.F., Murphy, D., 2007. Dehydration-induced proteome changes in the rat hypothalamo-neurohypophyseal system. Endocrinology 148, 3041–3052.
Greenwood, M., Bordieri, L., Greenwood, M.P., Rosso Melo, M., Colombari, D.S., Colombari, E., Paton, J.F., Murphy, D., 2014. Transcription factor CREB3L1 regulates vasopressin gene expression in the rat hypothalamus. J. Neurosci. 34, 3810–3820.
Greenwood, M., Greenwood, M.P., Mecawi, A.S., Loh, S.Y., Rodrigues, J.A., Paton, J.F., Murphy, D., 2015a. Transcription factor CREB3L1 mediates cAMP and glucocorticoid regulation of arginine vasopressin gene transcription in the rat hypothalamus. Mol. Brain 8, 68.
Greenwood, M., Greenwood, M.P., Paton, J.F., Murphy, D., 2015b. Transcription Factor CREB3L1 Regulates Endoplasmic Reticulum Stress Response Genes in the Osmotically Challenged Rat Hypothalamus. PLoS One 10, e0124956.
Greenwood, M.P., Greenwood, M., Gillard, B.T., Loh, S.Y., Paton, J.F., Murphy, D., 2016a. Epigenetic control of the vasopressin promoter explains physiological ability to regulate vasopressin transcription in dehydration and salt loading states in the rat. J. Neuroendocrinol. 28.
Greenwood, M.P., Greenwood, M., Mecawi, A.S., Antunes-Rodrigues, J., Paton, J.F., Murphy, D., 2016b. Rasd1, a small G protein with a big role in the hypothalamic response to neuronal activation. Mol. Brain 9, 1.
Guo, J.U., Su, Y., Zhong, C., Ming, G.L., Song, H., 2011. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell 145, 423–434.
Hatton, G.I., Walters, J.K., 1973. Induced multiple nucleoli, nucleolar margination, and cell size changes in supraoptic neurons during dehydration and rehydration in the rat. Brain Res. 59, 137–154.
Herman, J.P., Schafer, M.K., Watson, S.J., Sherman, T.G., 1991. In situ hybridization analysis of arginine vasopressin gene transcription using intron-specific probes. Mol. Endocrinol. 5, 1447–1456.
Hindmarch, C., Yao, S., Beighton, G., Paton, J., Murphy, D., 2006. A comprehensive description of the transcriptome of the hypothalamoneurohypophyseal system in euhydrated and dehydrated rats. Proc. Natl. Acad. Sci. U. S. A. 103, 1609–1614.
Husain, M.K., Fernando, N., Shapiro, M., Kagan, A., Glick, S.M., 1973. Radioimmunoassay of arginine vasopressin in human plasma. J. Clin. Endocrinol. Metab. 37, 616–625.
Ishunina, T.A., Swaab, D.F., 2002. Neurohypophyseal peptides in aging and Alzheimer's disease. Ageing Res. Rev. 1, 537–558.
Iwasaki, Y., Oiso, Y., Saito, H., Majzoub, J.A., 1997. Positive and negative regulation of the rat vasopressin gene promoter. Endocrinology 138, 5266–5274.
Keck, M.E., Hatzinger, M., Wotjak, C.T., Landgraf, R., Holsboer, F., Neumann, I.D., 2000. Ageing alters intrahypothalamic release patterns of vasopressin and oxytocin in rats. Eur. J. Neurosci. 12, 1487–1494.
Kmiec, Z., 2006. Central regulation of food intake in ageing. J Physiol. Pharmacol. 57 Suppl 6, 7–16.
Kondo, N., Arima, H., Banno, R., Kuwahara, S., Sato, I., Oiso, Y., 2004. Osmoregulation of vasopressin release and gene transcription under acute and chronic hypovolemia in rats. Am. J. Physiol. Endocrinol. Metab. 286, E337–E346.
Konopacka, A., Greenwood, M., Loh, S.Y., Paton, J., Murphy, D., 2015a. RNA binding protein Caprin-2 is a pivotal regulator of the central osmotic defense response. Elife 4.
Konopacka, A., Qiu, J., Yao, S.T., Greenwood, M.P., Greenwood, M., Lancaster, T., Inoue, W., Mecawi, A.S., Vechiato, F.M., de Lima, J.B., Coletti, R., Hoe, S.Z., Martin, A., Lee, J., Joseph, M., Hindmarch, C., Paton, J., Antunes-Rodrigues, J., Bains, J., Murphy, D., 2015b. Osmoregulation requires brain expression of the renal Na-K-2Cl cotransporter NKCC2. J. Neurosci. 35, 5144–5155.
Kruskal, W.H., Wallis, W.A., 1952. Use of ranks in one-criterion variance analysis. J. Am. Stat. Assoc. 47, 588–621.
Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25, 402–408.
Lucassen, P.J., Van Heerikhuize, J.J., Guldenaar, S.E., Pool, C.W., Hofman, M.A., Swaab, D.F., 1997. Unchanged amounts of vasopressin mRNA in the supraoptic and paraventricular nucleus during aging and in Alzheimer's disease. J. Neuroendocrinol. 9, 297–305.
Mack, G.W., Weseman, C.A., Langhans, G.W., Scherzer, H., Gillen, C.M., Nadel, E.R., 1994. Body fluid balance in dehydrated healthy older men: thirst and renal osmoregulation. J. Appl. Physiol. 76, 1615–1623.
McKinley, M.J., Mathai, M.L., McAllen, R.M., McClear, R.C., Miselis, R.R., Pennington, G.L., Vivas, L., Wade, J.D., Oldfield, B.J., 2004. Vasopressin secretion: osmotic and hormonal regulation by the lamina terminalis. J. Neuroendocrinol. 16, 340–347.
McLean, K.A., O'Neill, P.A., Davies, I., Morris, J., 1992. Influence of age on plasma osmolality: a community study. Age Aging 21, 56–60.
Mecawi Ade, S., Ruginsk, S.G., Elias, L.L., Varanda, W.A., Antunes-Rodrigues, J., 2015. Neuroendocrine regulation of hydromineral homeostasis. Compr. Physiol. 5, 1465–1516.
Murphy, D., Carter, D., 1990. Vasopressin gene expression in the rodent hypothalamus: transcriptional and posttranscriptional responses to physiological stimulation. Mol. Endocrinol. 4, 1051–1059.
Murphy, D., Konopacka, A., Hindmarch, C., Paton, J.F., Sweedler, J.V., Gillette, M.U., Ueta, Y., Grinevich, V., Lozic, M., Japundzic-Zigon, N., 2012. The hypothalamic-neurohypophyseal system: from genome to physiology. J. Neuroendocrinol. 24, 539–553.
Mzhavia, N., Berman, Y., Che, F.Y., Fricker, L.D., Devi, L.A., 2001. ProSAAS processing in mouse brain and pituitary. J. Biol. Chem. 276, 6207–6213.
Nissen, R., Hu, B., Renaud, L.P., 1994. N-methyl-D-aspartate receptor antagonist ketamine selectively attenuates spontaneous phasic activity of supraoptic vasopressin neurons in vivo. Neuroscience 59, 115–120.
Palatnik, C.M., Wilkins, C., Jacobson, A., 1984. Translational control during early Dictyostelium development: possible involvement of poly(A) sequences. Cell 36, 1017–1025.
Palin, K., Moreau, M.L., Orcel, H., Duvoid-Guillou, A., Rabie, A., Kelley, K.W., Moos, F., 2009. Age-impaired fluid homeostasis depends on the balance of IL-6/IGF-I in the rat supraoptic nuclei. Neurobiol. Aging 30, 1677–1692.
Pardy, K., Carter, D., Murphy, D., 1990. Dopaminergic mediation of physiological changes in proopiomelanocortin messenger ribonucleic acid expression in the neurointermediate lobe of the rat pituitary. Endocrinology 126, 2960–2964.
Phillips, P.A., Johnston, C.I., Gray, L., 1993. Disturbed fluid and electrolyte homoeostasis following dehydration in elderly people. Age Aging 22, S26–S33.
Pollack, J.A., Skvorak, J.P., Nazian, S.J., Landon, C.S., Dietz, J.R., 1997. Alterations in atrial natriuretic peptide (ANP) secretion and renal effects in aging. J. Gerontol. A. Biol. Sci. Med. Sci. 52, B196–B202.
Preisser, L., Houot, L., Teillet, L., Kortulewski, T., Morel, A., Tronik-Le Roux, D., Corman, B., 2004. Gene expression in aging kidney and pituitary. Biogerontology 5, 39–47.
Qiu, J., Yao, S., Hindmarch, C., Antunes, V., Paton, J., Murphy, D., 2007. Transcription factor expression in the hypothalamo-neurohypophyseal system of the dehydrated rat: upregulation of gonadotrophin inducible transcription factor 1 mRNA is mediated by cAMP-dependent protein kinase A. J. Neurosci. 27, 2196–2203.
Romanova, E.V., Aerts, J.T., Croushore, C.A., Sweedler, J.V., 2014. Small-volume analysis of cell-cell signaling molecules in the brain. Neuropsychopharmacology 39, 50–64.
Romanova, E.V., Dowd, S.E., Sweedler, J.V., 2013. Quantitation of endogenous peptides using mass spectrometry based methods. Curr. Opin. Chem. Biol. 17, 801–808.
Romanova, E.V., Rubakhin, S.S., Sweedler, J.V., 2008. One-step sampling, extraction, and storage protocol for peptidomics using dihydroxybenzoic Acid. Anal. Chem. 80, 3379–3386.
Romanova, E.V., Sweedler, J.V., 2015. Peptidomics for the discovery and characterization of neuropeptides and hormones. Trends Pharmacol. Sci. 36, 579–586.
Silver, A.J., Morley, J.E., Ishimaru-Tseng, T.V., Morley, P.M., 1993. Angiotensin II and fluid ingestion in old rats. Neurobiol. Aging 14, 519–522.
Silverman, W.F., Aravich, P.A., Sladek Jr., J.R., Sladek, C.D., 1990. Physiological and biochemical indices of neurohypophyseal function in the aging Fischer rat. Neuroendocrinology 52, 181–190.
Sladek, C.D., McNeill, T.H., Gregg, C.M., Blair, M.L., Baggs, R.B., 1981. Vasopressin and renin response to dehydration in aged rats. Neurobiol. Aging 2, 293–302.
Sladek, C.D., Olschowka, J.A., 1994. Dehydration induces Fos, but not increased vasopressin mRNA in the supraoptic nucleus of aged rats. Brain Res. 652, 207–215.
Snyder, N.A., Feigal, D.W., Arieff, A.I., 1987. Hypernatremia in elderly patients. A heterogeneous, morbid, and iatrogenic entity. Ann. Intern. Med. 107, 309–319.
Stephens, M.A., 1974. EDF for goodness of fit and some comparisons. J. Am. Stat. Assoc. 69, 730–737.
Swaab, D.F., Bao, A.M., 2011. (Re-)activation of neurons in aging and dementia: lessons from the hypothalamus. Exp. Gerontol. 46, 178–184.
Swenson, K.L., Sands, J.M., Jacobs, J.D., Sladek, C.D., 1997. Effect of aging on vasopressin and aquaporin responses to dehydration in Fischer 344-brown-Norway F1 rats. Am. J. Physiol. 273 (1 Pt 2), R35–R40.
Terwel, D., Markerink, M., Jolles, J., 1992. Age-related changes in concentrations of vasopressin in the central nervous system and plasma of the male Wistar rat. Mech. Aging Dev. 65, 127–136.
Verbalis, J.G., Mangione, M.P., Stricker, E.M., 1991. Oxytocin produces natriuresis in rats at physiological plasma concentrations. Endocrinology 128, 1317–1322.
Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics 1, 80–83.
Zbuzek, V., Fuchs, A.R., Zbuzek, V.K., Wu, W.H., 1988. Neurohypophyseal aging: differential changes in oxytocin and vasopressin release, studied in Fischer 344 and Sprague-Dawley rats. Neuroendocrinology 48, 619–626.
Zeevi, M., Nevins, J.R., Darnell Jr., J.E., 1982. Newly formed mRNA lacking polyadenylic acid enters the cytoplasm and the polyribosomes but has a shorter half-life in the absence of polyadenylic acid. Mol. Cell Biol. 2, 517–525.
Zhang, B., Glasgow, E., Murase, T., Verbalis, J.G., Gainer, H., 2001. Chronic hypoosmolality induces a selective decrease in magnocellular neurone soma and nuclear size in the rat hypothalamic supraoptic nucleus. J. Neuroendocrinol. 13, 29–36.
Zhang, X., Odom, D.T., Koo, S.H., Conkright, M.D., Canettieri, G., Best, J., Chen, H., Jenner, R., Herbolsheimer, E., Jacobsen, E., Kadam, S., Ecker, J.R., Emerson, B., Hogenesch, J.B., Unterman, T., Young, R.A., Montminy, M., 2005. Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues. Proc. Natl. Acad. Sci. U. S. A. 102, 4459–4464.
Zhang, Z., Bourque, C.W., 2003. Osmometry in osmosensory neurons. Nat. Neurosci. 6, 1021–1022.

Bibliography

[1] Tsung Fei Khang and Ching Yee Lau. Getting the most out of RNA-seq data analysis. *PeerJ*, 3:e1360, 2015.

[2] Stephen Hales. Statistical Essays: Containing Haemastaticks. *Innys, Manby and Woodward*, 1733.

[3] Mojtaba Heydari, Behnam Dalfardi, Samad E J Golzari, Hamzeh Habibi, and Mohammad Mehdi Zarshenas. The medieval origins of the concept of hypertension. *Heart views : the official journal of the Gulf Heart Association*, 15(3):96–98, jul 2014.

[4] P M Esunge. From blood pressure to hypertension: the history of research. *Journal of the Royal Society of Medicine*, 84(10):621, oct 1991.

[5] Scipione Riva-Rocci. Un nuovo sfigmomanometro. *Gazzetta medica di Torino*, 47:981–1001, 1896.

[6] Y L Shevchenko and J E Tsitlik. 90th Anniversary of the development by Nikolai S. Korotkoff of the auscultatory method of measuring blood pressure., jul 1996.

[7] THEODORE C JANEWAY. A CLINICAL STUDY OF HYPERTENSIVE CARDIOVASCULAR DISEASE. *JAMA Internal Medicine*, XII(6):755–798, 1913.

[8] Bryan Williams, Giuseppe Mancia, Wilko Spiering, Enrico Agabiti Rosei, Michel Azizi, Michel Burnier, Denis L Clement, Antonio Coca, Giovanni de Simone, Anna Dominiczak, Thomas Kahan, Felix Mahfoud, Josep Redon, Luis Ruilope, Alberto Zanchetti, Mary Kerins, Sverre E Kjeldsen, Reinhold Kreutz, Stephane Laurent, Gregory Y H Lip, Richard McManus, Krzysztof Narkiewicz, Frank Ruschitzka, Roland E Schmieder, Evgeny Shlyakhto, Costas Tsioufis, Victor Aboyans, and Ileana Desormais. 2018 ESC/ESH Guidelines for the management of arterial hypertension: The Task Force for the management of arterial hypertension of the European Society of Cardiology and the European Society of Hypertension: The Task Force for the management of arterial h. *Journal of hypertension*, 36(10):1953–2041, oct 2018.

[9] Paul K Whelton, Robert M Carey, Wilbert S Aronow, Donald E Jr Casey, Karen J Collins, Cheryl Dennison Himmelfarb, Sondra M DePalma, Samuel Gidding, Kenneth A Jamerson, Daniel W Jones, Eric J MacLaughlin, Paul Muntner, Bruce Ovbiagele, Sidney C Jr Smith, Crystal C Spencer, Randall S Stafford, Sandra J Taler, Randal J Thomas, Kim A Sr Williams, Jeff D Williamson, and Jackson T Jr Wright. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline

for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task F. *Hypertension (Dallas, Tex. : 1979)*, 71(6):1269–1324, jun 2018.

[10] A Randomized Trial of Intensive versus Standard Blood-Pressure Control. *New England Journal of Medicine*, 373(22):2103–2116, 2015.

[11] Aram V Chobanian, George L Bakris, Henry R Black, William C Cushman, Lee A Green, Joseph L Jr Izzo, Daniel W Jones, Barry J Materson, Suzanne Oparil, Jackson T Jr Wright, and Edward J Roccella. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. *JAMA*, 289(19):2560–2572, may 2003.

[12] Sarah Lewington, Robert Clarke, Nawab Qizilbash, Richard Peto, and Rory Collins. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet (London, England)*, 360(9349):1903–1913, dec 2002.

[13] C John Dickinson. Strokes and their relationship to hypertension. *Current opinion in nephrology and hypertension*, 12(1):91–96, jan 2003.

[14] D R J Singer and A Kite. Management of hypertension in peripheral arterial disease: does the choice of drugs matter? *European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery*, 35(6):701–708, jun 2008.

[15] Dennis H Lau, Stanley Nattel, Jonathan M Kalman, and Prashanthan Sanders. Modifiable Risk Factors and Atrial Fibrillation. *Circulation*, 136(6):583–596, aug 2017.

[16] Richard E Tracy and Steven White. A method for quantifying adrenocortical nodular hyperplasia at autopsy: some use of the method in illuminating hypertension and atherosclerosis. *Annals of diagnostic pathology*, 6(1):20–29, feb 2002.

[17] Christophe Tzourio. Hypertension, cognitive decline, and dementia: an epidemiological perspective. *Dialogues in clinical neuroscience*, 9(1):61–70, 2007.

[18] Alicia J Jenkins, Mugdha V Joglekar, Anandwardhan A Hardikar, Anthony C Keech, David N O'Neal, and Andrzej S Januszewski. Biomarkers in Diabetic Retinopathy. *The review of diabetic studies : RDS*, 12(1-2):159–195, 2015.

[19] Global Heath Risks: Mortality and burden of disease attributable to selected major risks. *World Health Organization*, 2009.

[20] World Health Organization. Causes of death 2008 : data sources and methods, 2011.

[21] World Health Organization. A global brief on Hypertension - World Health Day 2013, 2013.

[22] Patricia M Kearney, Megan Whelton, Kristi Reynolds, Paul Muntner, Paul K Whelton, and Jiang He. Global burden of hypertension: analysis of worldwide data. *Lancet (London, England)*, 365(9455):217–223, jan 2005.

[23] Thomas A Gaziano, Asaf Bitton, Shuchi Anand, and Milton C Weinstein. The global cost of nonoptimal blood pressure. *Journal of hypertension*, 27(7):1472–1477, jul 2009.

[24] R PLATT. Heredity in hypertension. *The Quarterly journal of medicine*, 16(3):111–133, jul 1947.

[25] GW Pickering. The NATURE of essential hypertension. *Lancet (London, England)*, 2(7110):1027–1030, dec 1959.

[26] Georg B Ehret, Patricia B Munroe, Kenneth M Rice, Murielle Bochud, Andrew D Johnson, Daniel I Chasman, Albert V Smith, Martin D Tobin, Germaine C Verwoert, Shih-Jen Hwang, Vasyl Pihur, Peter Vollenweider, Paul F O'Reilly, Najaf Amin, Jennifer L Bragg-Gresham, Alexander Teumer, Nicole L Glazer, Lenore Launer, Jing Hua Zhao, Yurii Aulchenko, Simon Heath, Siim Sober, Afshin Parsa, Jian'an Luan, Pankaj Arora, Abbas Dehghan, Feng Zhang, Gavin Lucas, Andrew A Hicks, Anne U Jackson, John F Peden, Toshiko Tanaka, Sarah H Wild, Igor Rudan, Wilmar Igl, Yuri Milaneschi, Alex N Parker, Cristiano Fava, John C Chambers, Ervin R Fox, Meena Kumari, Min Jin Go, Pim van der Harst, Wen Hong Linda Kao, Marketa Sjogren, D G Vinay, Myriam Alexander, Yasuharu Tabara, Sue Shaw-Hawkins, Peter H Whincup, Yongmei Liu, Gang Shi, Johanna Kuusisto, Bamidele Tayo, Mark Seielstad, Xueling Sim, Khanh-Dung Hoang Nguyen, Terho Lehtimaki, Giuseppe Matullo, Ying Wu, Tom R Gaunt, N Charlotte Onland-Moret, Matthew N Cooper, Carl G P Platou, Elin Org, Rebecca Hardy, Santosh Dahgam, Jutta Palmen, Veronique Vitart, Peter S Braund, Tatiana Kuznetsova, Cuno S P M Uiterwaal, Adebowale Adeyemo, Walter Palmas, Harry Campbell, Barbara Ludwig, Maciej Tomaszewski, Ioanna Tzoulaki, Nicholette D Palmer, Thor Aspelund, Melissa Garcia, Yen-Pei C Chang, Jeffrey R O'Connell, Nanette I Steinle, Diederick E Grobbee, Dan E Arking, Sharon L Kardia, Alanna C Morrison, Dena Hernandez, Samer Najjar, Wendy L McArdle, David Hadley, Morris J Brown, John M Connell, Aroon D Hingorani, Ian N M Day, Debbie A Lawlor, John P Beilby, Robert W Lawrence, Robert Clarke, Jemma C Hopewell, Halit Ongen, Albert W Dreisbach, Yali Li, J Hunter Young, Joshua C Bis, Mika Kahonen, Jorma Viikari, Linda S Adair, Nanette R Lee, Ming-Huei Chen, Matthias Olden, Cristian Pattaro, Judith A Hoffman Bolton, Anna Kottgen, Sven Bergmann, Vincent Mooser, Nish Chaturvedi, Timothy M Frayling, Muhammad Islam, Tazeen H Jafar, Jeanette Erdmann, Smita R Kulkarni, Stefan R Bornstein, Jurgen Grassler, Leif Groop, Benjamin F Voight, Johannes Kettunen, Philip Howard, Andrew Taylor, Simonetta Guarrera, Fulvio Ricceri, Valur Emilsson, Andrew Plump, Ines Barroso, Kay-Tee Khaw, Alan B Weder, Steven C Hunt, Yan V Sun, Richard N Bergman, Francis S Collins, Lori L Bonnycastle, Laura J Scott, Heather M Stringham, Leena Peltonen, Markus Perola, Erkki Vartiainen, Stefan-Martin Brand, Jan A Staessen, Thomas J Wang, Paul R Burton, Maria Soler Artigas, Yanbin Dong, Harold Snieder, Xiaoling Wang, Haidong

Zhu, Kurt K Lohman, Megan E Rudock, Susan R Heckbert, Nicholas L Smith, Kerri L Wiggins, Ayo Doumatey, Daniel Shriner, Gudrun Veldre, Margus Viigimaa, Sanjay Kinra, Dorairaj Prabhakaran, Vikal Tripathy, Carl D Langefeld, Annika Rosengren, Dag S Thelle, Anna Maria Corsi, Andrew Singleton, Terrence Forrester, Gina Hilton, Colin A McKenzie, Tunde Salako, Naoharu Iwai, Yoshikuni Kita, Toshio Ogihara, Takayoshi Ohkubo, Tomonori Okamura, Hirotsugu Ueshima, Satoshi Umemura, Susana Eyheramendy, Thomas Meitinger, H-Erich Wichmann, Yoon Shin Cho, Hyung-Lae Kim, Jong-Young Lee, James Scott, Joban S Sehmi, Weihua Zhang, Bo Hedblad, Peter Nilsson, George Davey Smith, Andrew Wong, Narisu Narisu, Alena Stancakova, Leslie J Raffel, Jie Yao, Sekar Kathiresan, Christopher J O'Donnell, Stephen M Schwartz, M Arfan Ikram, W T Jr Longstreth, Thomas H Mosley, Sudha Seshadri, Nick R G Shrine, Louise V Wain, Mario A Morken, Amy J Swift, Jaana Laitinen, Inga Prokopenko, Paavo Zitting, Jackie A Cooper, Steve E Humphries, John Danesh, Asif Rasheed, Anuj Goel, Anders Hamsten, Hugh Watkins, Stephan J L Bakker, Wiek H van Gilst, Charles S Janipalli, K Radha Mani, Chittaranjan S Yajnik, Albert Hofman, Francesco U S Mattace-Raso, Ben A Oostra, Ayse Demirkan, Aaron Isaacs, Fernando Rivadeneira, Edward G Lakatta, Marco Orru, Angelo Scuteri, Mika Ala-Korpela, Antti J Kangas, Leo-Pekka Lyytikainen, Pasi Soininen, Taru Tukiainen, Peter Wurtz, Rick Twee-Hee Ong, Marcus Dorr, Heyo K Kroemer, Uwe Volker, Henry Volzke, Pilar Galan, Serge Hercberg, Mark Lathrop, Diana Zelenika, Panos Deloukas, Massimo Mangino, Tim D Spector, Guangju Zhai, James F Meschia, Michael A Nalls, Pankaj Sharma, Janos Terzic, M V Kranthi Kumar, Matthew Denniff, Ewa Zukowska-Szczechowska, Lynne E Wagenknecht, F Gerald R Fowkes, Fadi J Charchar, Peter E H Schwarz, Caroline Hayward, Xiuqing Guo, Charles Rotimi, Michiel L Bots, Eva Brand, Nilesh J Samani, Ozren Polasek, Philippa J Talmud, Fredrik Nyberg, Diana Kuh, Maris Laan, Kristian Hveem, Lyle J Palmer, Yvonne T van der Schouw, Juan P Casas, Karen L Mohlke, Paolo Vineis, Olli Raitakari, Santhi K Ganesh, Tien Y Wong, E Shyong Tai, Richard S Cooper, Markku Laakso, Dabeeru C Rao, Tamara B Harris, Richard W Morris, Anna F Dominiczak, Mika Kivimaki, Michael G Marmot, Tetsuro Miki, Danish Saleheen, Giriraj R Chandak, Josef Coresh, Gerjan Navis, Veikko Salomaa, Bok-Ghee Han, Xiaofeng Zhu, Jaspal S Kooner, Olle Melander, Paul M Ridker, Stefania Bandinelli, Ulf B Gyllensten, Alan F Wright, James F Wilson, Luigi Ferrucci, Martin Farrall, Jaakko Tuomilehto, Peter P Pramstaller, Roberto Elosua, Nicole Soranzo, Eric J G Sijbrands, David Altshuler, Ruth J F Loos, Alan R Shuldiner, Christian Gieger, Pierre Meneton, Andre G Uitterlinden, Nicholas J Wareham, Vilmundur Gudnason, Jerome I Rotter, Rainer Rettig, Manuela Uda, David P Strachan, Jacqueline C M Witteman, Anna-Liisa Hartikainen, Jacques S Beckmann, Eric Boerwinkle, Ramachandran S Vasan, Michael Boehnke, Martin G Larson, Marjo-Riitta Jarvelin, Bruce M Psaty, Goncalo R Abecasis, Aravinda Chakravarti, Paul Elliott, Cornelia M van Duijn, Christopher Newton-Cheh, Daniel Levy, Mark J Caulfield, and Toby Johnson. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*, 478(7367):103–109, sep 2011.

[27] Norihiro Kato, Fumihiko Takeuchi, Yasuharu Tabara, Tanika N Kelly, Min Jin Go, Xueling Sim, Wan Ting Tay, Chien-Hsiun Chen, Yi Zhang, Ken Yamamoto,

Tomohiro Katsuya, Mitsuhiro Yokota, Young Jin Kim, Rick Twee Hee Ong, Toru Nabika, Dongfeng Gu, Li-Ching Chang, Yoshihiro Kokubo, Wei Huang, Keizo Ohnaka, Yukio Yamori, Eitaro Nakashima, Cashell E Jaquish, Jong-Young Lee, Mark Seielstad, Masato Isono, James E Hixson, Yuan-Tsong Chen, Tetsuro Miki, Xueya Zhou, Takao Sugiyama, Jae-Pil Jeon, Jian Jun Liu, Ryoichi Takayanagi, Sung Soo Kim, Tin Aung, Yun Ju Sung, Xuegong Zhang, Tien Yin Wong, Bok-Ghee Han, Shotai Kobayashi, Toshio Ogihara, Dingliang Zhu, Naoharu Iwai, Jer-Yuarn Wu, Yik Ying Teo, E Shyong Tai, Yoon Shin Cho, and Jiang He. Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nature genetics*, 43(6):531–538, jun 2011.

[28] Daniel Levy, Georg B Ehret, Kenneth Rice, Germaine C Verwoert, Lenore J Launer, Abbas Dehghan, Nicole L Glazer, Alanna C Morrison, Andrew D Johnson, Thor Aspelund, Yurii Aulchenko, Thomas Lumley, Anna Kottgen, Ramachandran S Vasan, Fernando Rivadeneira, Gudny Eiriksdottir, Xiuqing Guo, Dan E Arking, Gary F Mitchell, Francesco U S Mattace-Raso, Albert V Smith, Kent Taylor, Robert B Scharpf, Shih-Jen Hwang, Eric J G Sijbrands, Joshua Bis, Tamara B Harris, Santhi K Ganesh, Christopher J O'Donnell, Albert Hofman, Jerome I Rotter, Josef Coresh, Emelia J Benjamin, Andre G Uitterlinden, Gerardo Heiss, Caroline S Fox, Jacqueline C M Witteman, Eric Boerwinkle, Thomas J Wang, Vilmundur Gudnason, Martin G Larson, Aravinda Chakravarti, Bruce M Psaty, and Cornelia M van Duijn. Genome-wide association study of blood pressure and hypertension. *Nature genetics*, 41(6):677–687, jun 2009.

[29] R P Lifton, A G Gharavi, and D S Geller. Molecular mechanisms of human hypertension. *Cell*, 104(4):545–556, feb 2001.

[30] Sandosh Padmanabhan, Christopher Newton-Cheh, and Anna F Dominiczak. Genetic basis of blood pressure and hypertension. *Trends in Genetics*, 28(8):397–408, 2012.

[31] Masao Omura, Jun Saito, Kunio Yamaguchi, Yukio Kakuta, and Tetsuo Nishikawa. Prospective study on the prevalence of secondary hypertension among hypertensive patients visiting a general outpatient clinic in Japan. *Hypertension research : official journal of the Japanese Society of Hypertension*, 27(3):193–202, mar 2004.

[32] Wanpen Vongpatanasin. Resistant hypertension: a review of diagnosis and management. *JAMA*, 311(21):2216–2224, jun 2014.

[33] Allen W Cowley. The genetic dissection of essential hypertension. *Nature Reviews Genetics*, 7(11):829–840, 2006.

[34] Jouke-Jan Hottenga, Dorret I Boomsma, Nina Kupper, Danielle Posthuma, Harold Snieder, Gonneke Willemsen, and Eco J C de Geus. Heritability and Stability of Resting Blood Pressure. *Twin Research and Human Genetics*, 8(5):499–508, 2005.

[35] Nina Kupper, Gonneke Willemsen, Harriette Riese, Danielle Posthuma, Dorret I Boomsma, and Eco J C de Geus. Heritability of daytime ambulatory blood pressure in an extended twin design. *Hypertension (Dallas, Tex. : 1979)*, 45(1):80–85, jan 2005.

[36] Somasundaram Arumugam, Remya Sreedhar, Rajarajan A Thandavarayan, Vengadeshprabhu Karuppagounder, Prasanna Krishnamurthy, Kenji Suzuki, Masahiko Nakamura, and Kenichi Watanabe. Angiotensin receptor blockers: Focus on cardiac and renal injury. *Trends in cardiovascular medicine*, 26(3):221–228, apr 2016.

[37] P Primatesta, M Brookes, and N R Poulter. Improved hypertension management and control: results from the health survey for England 1998. *Hypertension (Dallas, Tex. : 1979)*, 38(4):827–832, oct 2001.

[38] Samuel J Mann. Neurogenic essential hypertension revisited: the case for increased clinical and research attention. *American journal of hypertension*, 16(10):881–888, oct 2003.

[39] Sandeep Chopra, Chris Baby, and Jubbin Jagan Jacob. Neuro-endocrine regulation of blood pressure. *Indian journal of endocrinology and metabolism*, 15 Suppl 4(Suppl4):S281–8, oct 2011.

[40] Charles Colin Thomas Hindmarch and David Murphy. The transcriptome and the hypothalamo-neurohypophyseal system. *Endocrine development*, 17:1–10, 2010.

[41] E Badoer. Hypothalamic paraventricular nucleus and cardiovascular regulation. *Clinical and experimental pharmacology & physiology*, 28(1-2):95–99, 2001.

[42] L V Scott and T G Dinan. Vasopressin and the regulation of hypothalamic-pituitary-adrenal axis function: implications for the pathophysiology of depression. *Life sciences*, 62(22):1985–1998, 1998.

[43] Yi-Chun Loraine Tung, Marcella Ma, Sarah Piper, Anthony Coll, Stephen O'Rahilly, and Giles S H Yeo. Novel leptin-regulated genes revealed by transcriptional profiling of the hypothalamic paraventricular nucleus. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 28(47):12419–12426, nov 2008.

[44] P Wiesel, L Mazzolai, J Nussberger, and T Pedrazzini. Two-kidney, one clip and one-kidney, one clip hypertension in mice. *Hypertension (Dallas, Tex. : 1979)*, 29(4):1025–1030, apr 1997.

[45] K Okamoto and K Aoki. Development of a strain of spontaneously hypertensive rats. *Japanese circulation journal*, 27:282–93, mar 1963.

[46] Y M Pinto, M Paul, and D Ganten. Lessons from rat models of hypertension: from Goldblatt to genetic engineering. *Cardiovascular research*, 39(1):77–88, jul 1998.

[47] C H Conrad, W W Brooks, K G Robinson, and O H Bing. Impaired myocardial function in spontaneously hypertensive rats with heart failure. *The American journal of physiology*, 260(1 Pt 2):H136–45, jan 1991.

[48] J M Pfeffer, M A Pfeffer, M C Fishbein, and E D Frohlich. Cardiac function and morphology with aging in the spontaneously hypertensive rat. *The American journal of physiology*, 237(4):H461–8, oct 1979.

[49] T F Luscher, L L Aarhus, and P M Vanhoutte. Indomethacin improves the impaired endothelium-dependent relaxations in small mesenteric arteries of the spontaneously hypertensive rat. *American journal of hypertension*, 3(1):55–58, jan 1990.

[50] T Wada, T Sanada, M Ojima, R Kanagawa, K Nishikawa, and Y Inada. Combined effects of the angiotensin II antagonist candesartan cilexetil (TCV-116) and other classes of antihypertensive drugs in spontaneously hypertensive rats. *Hypertension research : official journal of the Japanese Society of Hypertension*, 19(4):247–254, dec 1996.

[51] T Takenaka, M Asano, K Shiono, M Shibasaki, and O Inagaki. Cardiovascular pharmacology of nicardipine in animals. *British journal of clinical pharmacology*, 20 Suppl 1:7S–22S, 1985.

[52] C Limas, B Westrum, and C J Limas. Comparative effects of hydralazine and captopril on the cardiovascular changes in spontaneously hypertensive rats. *The American journal of pathology*, 117(3):360–371, dec 1984.

[53] H Karam, D Heudes, P Bruneval, M F Gonzales, B M Loffler, M Clozel, and J P Clozel. Endothelin antagonism in end-organ damage of spontaneously hypertensive rats. Comparison with angiotensin-converting enzyme inhibition and calcium antagonism. *Hypertension (Dallas, Tex. : 1979)*, 28(3):379–385, sep 1996.

[54] J Stamler. The INTERSALT Study: background, methods, findings, and implications. *The American journal of clinical nutrition*, 65(2 Suppl):626S–642S, feb 1997.

[55] Sung Kyu Ha. Dietary salt intake and hypertension. *Electrolyte & blood pressure : E & BP*, 12(1):7–18, jun 2014.

[56] G R MENEELY and C O BALL. Experimental epidemiology of chronic sodium chloride toxicity and the protective effect of potassium chloride. *The American journal of medicine*, 25(5):713–725, nov 1958.

[57] L K DAHL and M HEINE. Effects of chronic excess salt feeding. Enhanced hypertensogenic effect of sea salt over sodium chloride. *The Journal of experimental medicine*, 113:1067–1076, jun 1961.

[58] L K DAHL, M HEINE, and L TASSINARI. Effects of chronia excess salt ingestion. Evidence that genetic factors play an important role in susceptibility to experimental hypertension. *The Journal of experimental medicine*, 115:1173–1190, jun 1962.

[59] L K DAHL, M HEINE, and L TASSINARI. Role of genetic factors in susceptibility to experimental hypertension due to chronic excess salt ingestion. *Nature*, 194:480–482, may 1962.

[60] PHE. Cardiovascular disease key facts: 7 Hypertension. *Public Health England*, 2013.

[61] Roger A Sunde. mRNA transcripts as molecular biomarkers in medicine and nutrition. *The Journal of nutritional biochemistry*, 21(8):665–670, aug 2010.

[62] Brian G Bazzell, William E Rainey, Richard J Auchus, Davide Zocco, Marco Bruttini, Scott L Hummel, and James Brian Byrd. Human Urinary mRNA as a Biomarker of Cardiovascular Disease. *Circulation. Genomic and precision medicine*, 11(9):e002213, sep 2018.

[63] Emma E Laing, Carla S Möller-Levet, Derk-Jan Dijk, and Simon N Archer. Identifying and validating blood mRNA biomarkers for acute and chronic insufficient sleep in humans: a machine learning approach. *Sleep*, 42(1), 2018.

[64] F Sanger, S Nicklen, and A R Coulson. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12):5463–5467, dec 1977.

[65] Chandra Shekhar Pareek, Rafal Smoczynski, and Andrzej Tretyn. Sequencing technologies and genome sequencing. *Journal of applied genetics*, 52(4):413–435, nov 2011.

[66] NHGRI. Human Genome Project - FAQ.

[67] Simon Andrews. FastQC: a quality control tool for high throughput sequence data., 2010.

[68] P Richterich. Estimation of errors in "raw" DNA sequences: a validation study. *Genome research*, 8(3):251–259, mar 1998.

[69] EcSeq. Why does the per base sequence quality decrease over the read in Illumina?, 2017.

[70] Alexander Dobin, Carrie A Davis, Chris Zaleski, Felix Schlesinger, Jorg Drenkow, Mark Chaisson, Philippe Batut, Sonali Jha, and Thomas R Gingeras. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21, oct 2012.

[71] Yang Liao, Gordon K Smyth, and Wei Shi. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, 30(7):923–930, apr 2014.

[72] Simon Anders, Paul Theodor Pyl, and Wolfgang Huber. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)*, 31(2):166–169, jan 2015.

[73] Ali Mortazavi, Brian A Williams, Kenneth McCue, Lorian Schaeffer, and Barbara Wold. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods*, 5(7):621–628, jul 2008.

[74] Ana Conesa, Pedro Madrigal, Sonia Tarazona, David Gomez-Cabrero, Alejandra Cervera, Andrew McPherson, Michał Wojciech Szcześniak, Daniel J Gaffney, Laura L Elo, Xuegong Zhang, and Ali Mortazavi. A survey of best practices for RNA-seq data analysis. *Genome biology*, 17:13, jan 2016.

[75] James H Bullard, Elizabeth Purdom, Kasper D Hansen, and Sandrine Dudoit. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC bioinformatics*, 11:94, feb 2010.

[76] Kasper D Hansen, Steven E Brenner, and Sandrine Dudoit. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic acids research*, 38(12):e131, jul 2010.

[77] Mark D Robinson, Davis J McCarthy, and Gordon K Smyth. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, 26(1):139–140, jan 2010.

[78] Simon Anders and Wolfgang Huber. Differential expression analysis for sequence count data. *Genome biology*, 11(10):R106, 2010.

[79] Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12):550, 2014.

[80] Charlotte Soneson and Mauro Delorenzi. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, 14(1):91, mar 2013.

[81] Franck Rapaport, Raya Khanin, Yupu Liang, Mono Pirun, Azra Krek, Paul Zumbo, Christopher E Mason, Nicholas D Socci, and Doron Betel. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome biology*, 14(9):R95, 2013.

[82] Fatemeh Seyednasrollah, Asta Laiho, and Laura L Elo. Comparison of software packages for detecting differential expression in RNA-seq studies. *Briefings in bioinformatics*, 16(1):59–70, jan 2015.

[83] Amy S Leonardson, Jun Zhu, Yanqing Chen, Kai Wang, John R Lamb, Marc Reitman, Valur Emilsson, and Eric E Schadt. The effect of food intake on gene expression in human peripheral blood. *Human molecular genetics*, 19(1):159–169, jan 2010.

[84] Tanja Zeller, Philipp Wild, Silke Szymczak, Maxime Rotival, Arne Schillert, Raphaele Castagne, Seraya Maouche, Marine Germain, Karl Lackner, Heidi Rossmann, Medea Eleftheriadis, Christoph R Sinning, Renate B Schnabel, Edith Lubos, Detlev Mennerich, Werner Rust, Claire Perret, Carole Proust, Viviane Nicaud, Joseph Loscalzo, Norbert Hubner, David Tregouet, Thomas Munzel, Andreas Ziegler, Laurence Tiret, Stefan Blankenberg, and Francois Cambien. Genetics and beyond–the transcriptome of human monocytes and disease susceptibility. *PloS one*, 5(5):e10693, may 2010.

[85] Todd M Bull, Christopher D Coldren, Mark Moore, Sylk M Sotto-Santiago, David V Pham, S Patrick Nana-Sinkam, Norbert F Voelkel, and Mark W Geraci. Gene microarray analysis of peripheral blood cells in pulmonary arterial hypertension. *American journal of respiratory and critical care medicine*, 170(8):911–919, oct 2004.

[86] Melvin T Korkor, Fan Bo Meng, Shen Yang Xing, Mu Chun Zhang, Jin Rui Guo, Xiao Xue Zhu, and Ping Yang. Microarray analysis of differential gene expression profile in peripheral blood cells of patients with human essential hypertension. *International journal of medical sciences*, 8(2):168–179, feb 2011.

[87] Tianxiao Huan, Tõnu Esko, Marjolein J Peters, Luke C Pilling, Katharina Schramm, Claudia Schurmann, Brian H Chen, Chunyu Liu, Roby Joehanes, Andrew D Johnson, Chen Yao, Sai-xia Ying, Paul Courchesne, Lili Milani, Nalini Raghavachari, Richard Wang, Poching Liu, Eva Reinmaa, Abbas Dehghan, Albert Hofman, André G Uitterlinden, Dena G Hernandez, Stefania Bandinelli, Andrew Singleton, David Melzer, Andres Metspalu, Maren Carstensen, Harald Grallert, Christian Herder, Thomas Meitinger, Annette Peters, Michael Roden, Melanie Waldenberger, Marcus Dörr, Stephan B Felix, Tanja Zeller, International Consortium for Blood Pressure GWAS (ICBP), Ramachandran Vasan, Christopher J O'Donnell, Peter J Munson, Xia Yang, Holger Prokisch, Uwe Völker, Joyce B J van Meurs, Luigi Ferrucci, and Daniel Levy. A Meta-analysis of Gene Expression Signatures of Blood Pressure and Hypertension. *PLOS Genetics*, 11(3):1–29, 2015.

[88] Cole Trapnell, Lior Pachter, and Steven L Salzberg. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)*, 25(9):1105–1111, may 2009.

[89] Jian Ye, George Coulouris, Irena Zaretskaya, Ioana Cutcutache, Steve Rozen, and Thomas L Madden. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC bioinformatics*, 13:134, jun 2012.

[90] Frank E Grubbs. Procedures for Detecting Outlying Observations in Samples. *Technometrics*, 11(1):1–21, 1969.

[91] K J Livak and T D Schmittgen. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif.)*, 25(4):402–408, dec 2001.

[92] Jennifer R Smith, G Thomas Hayman, Shur-Jen Wang, Stanley J F Laulederkind, Matthew J Hoffman, Mary L Kaldunski, Monika Tutaj, Jyothi Thota, Harika S Nalabolu, Santoshi L R Ellanki, Marek A Tutaj, Jeffrey L De Pons, Anne E Kwitek, Melinda R Dwinell, and Mary E Shimoyama. The Year of the Rat: The Rat Genome Database at 20: a multi-species knowledgebase and analysis platform. *Nucleic acids research*, 48(D1):D731–D742, jan 2020.

[93] Babraham Bioinformatics. FastQC Help Modules.

[94] Mingkwan Greenwood, Michael Paul Greenwood, Julian F R Paton, and David Murphy. Transcription Factor CREB3L1 Regulates Endoplasmic Reticulum Stress Response Genes in the Osmotically Challenged Rat Hypothalamus. *PloS one*, 10(4):e0124956, 2015.

[95] Michael W Pfaffl, Ales Tichopad, Christian Prgomet, and Tanja P Neuvians. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. *Biotechnology letters*, 26(6):509–515, mar 2004.

[96] Shanrong Zhao, Ying Zhang, Ramya Gamini, Baohong Zhang, and David von Schack. Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: polyA+ selection versus rRNA depletion. *Scientific Reports*, 8(1):4781, 2018.

[97] Wei Pan, Jizhen Lin, and Chap T Le. How many replicates of arrays are required to detect gene expression changes in microarray experiments? A mixture model approach. *Genome biology*, 3(5):research0022, 2002.

[98] Gary A Churchill. Fundamentals of experimental design for cDNA microarrays. *Nature genetics*, 32 Suppl:490–495, dec 2002.

[99] Daniel Bottomly, Nicole A R Walter, Jessica Ezzell Hunter, Priscila Darakjian, Sunita Kawane, Kari J Buck, Robert P Searles, Michael Mooney, Shannon K McWeeney, and Robert Hitzemann. Evaluating gene expression in C57BL/6J and DBA/2J mouse striatum using RNA-Seq and microarrays. *PloS one*, 6(3):e17820, mar 2011.

[100] Conrad J Burden, Sumaira E Qureshi, and Susan R Wilson. Error estimates for the analysis of differential expression from RNA-seq count data. *PeerJ*, 2:e576, 2014.

[101] Nicholas J Schurch, Pieta Schofield, Marek Gierlinski, Christian Cole, Alexander Sherstnev, Vijender Singh, Nicola Wrobel, Karim Gharbi, Gordon G Simpson, Tom Owen-Hughes, Mark Blaxter, and Geoffrey J Barton. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA (New York, N.Y.)*, 22(6):839–851, jun 2016.

[102] Charlotte Soneson. compcodeR–an R package for benchmarking differential expression methods for RNA-seq data. *Bioinformatics (Oxford, England)*, 30(17):2517–2518, sep 2014.

[103] Dominic O'Neil, Heike Glowatz, and Martin Schlumpberger. Ribosomal RNA depletion for efficient use of RNA-seq capacity. *Current protocols in molecular biology*, Chapter 4:Unit 4.19, jul 2013.

[104] Ashwini Kumar, Matti Kankainen, Alun Parsons, Olli Kallioniemi, Pirkko Mattila, and Caroline A Heckman. The impact of RNA sequence library construction protocols on transcriptomic profiling of leukemia. *BMC genomics*, 18(1):629, aug 2017.

[105] Sven Schuierer, Walter Carbone, Judith Knehr, Virginie Petitjean, Anita Fernandez, Marc Sultan, and Guglielmo Roma. A comprehensive assessment of RNA-seq protocols for degraded and low-quantity samples. *BMC genomics*, 18(1):442, jun 2017.

[106] Yan Guo, Shilin Zhao, Quanhu Sheng, Mingsheng Guo, Brian Lehmann, Jennifer Pietenpol, David C Samuels, and Yu Shyr. RNAseq by Total RNA Library Identifies Additional RNAs Compared to Poly(A) RNA Library. *BioMed research international*, 2015:862130, 2015.

[107] Adriana Alberti, Caroline Belser, Stefan Engelen, Laurie Bertrand, Celine Orvain, Laura Brinas, Corinne Cruaud, Laurene Giraut, Corinne Da Silva, Cyril Firmo, Jean-Marc Aury, and Patrick Wincker. Comparison of library preparation methods reveals their impact on interpretation of metatranscriptomic data. *BMC genomics*, 15:912, oct 2014.

[108] W James Kent. BLAT–the BLAST-like alignment tool. *Genome research*, 12(4):656–664, apr 2002.

[109] Nuno A Fonseca, Johan Rung, Alvis Brazma, and John C Marioni. Tools for mapping high-throughput sequencing data. *Bioinformatics (Oxford, England)*, 28(24):3169–3177, dec 2012.

[110] Pär G Engström, Tamara Steijger, Botond Sipos, Gregory R Grant, André Kahles, The RGASP Consortium, Tyler Alioto, Jonas Behr, Paul Bertone, Regina Bohnert, Davide Campagna, Carrie A Davis, Alexander Dobin, Pär G Engström, Thomas R Gingeras, Nick Goldman, Gregory R Grant, Roderic Guigó, Jennifer Harrow, Tim J Hubbard, Géraldine Jean, André Kahles, Peter Kosarev, Sheng Li, Jinze Liu, Christopher E Mason, Vladimir Molodtsov, Zemin Ning, Hannes Ponstingl, Jan F Prins, Gunnar Rätsch, Paolo Ribeca, Igor Seledtsov, Botond Sipos, Victor Solovyev, Tamara Steijger, Giorgio Valle, Nicola Vitulo, Kai Wang, Thomas D Wu, Georg Zeller, Gunnar Rätsch, Nick Goldman, Tim J Hubbard, Jennifer Harrow, Roderic Guigó, and Paul Bertone. Systematic evaluation of spliced alignment programs for RNA-seq data. *Nature Methods*, 10:1185, nov 2013.

[111] Sarah Djebali, Carrie A Davis, Angelika Merkel, Alex Dobin, Timo Lassmann, Ali Mortazavi, Andrea Tanzer, Julien Lagarde, Wei Lin, Felix Schlesinger, Chenghai Xue, Georgi K Marinov, Jainab Khatun, Brian A Williams, Chris Zaleski, Joel Rozowsky, Maik Roder, Felix Kokocinski, Rehab F Abdelhamid, Tyler Alioto, Igor Antoshechkin, Michael T Baer, Nadav S Bar, Philippe Batut, Kimberly Bell, Ian Bell, Sudipto Chakrabortty, Xian Chen, Jacqueline Chrast, Joao Curado, Thomas Derrien, Jorg Drenkow, Erica Dumais, Jacqueline Dumais, Radha Duttagupta, Emilie Falconnet, Meagan Fastuca, Kata Fejes-Toth, Pedro Ferreira, Sylvain Foissac, Melissa J Fullwood, Hui Gao, David Gonzalez, Assaf Gordon, Harsha Gunawardena, Cedric Howald, Sonali Jha, Rory Johnson, Philipp Kapranov, Brandon King, Colin Kingswood, Oscar J Luo, Eddie Park, Kimberly Persaud, Jonathan B Preall, Paolo Ribeca, Brian Risk, Daniel Robyr, Michael Sammeth, Lorian Schaffer, Lei-Hoon See, Atif Shahab, Jorgen Skancke, Ana Maria Suzuki, Hazuki Takahashi, Hagen Tilgner, Diane Trout, Nathalie Walters, Huaien Wang, John Wrobel, Yanbao Yu, Xiaoan Ruan, Yoshihide Hayashizaki, Jennifer Harrow, Mark Gerstein, Tim Hubbard, Alexandre Reymond, Stylianos E Antonarakis, Gregory Hannon, Morgan C Giddings,

Yijun Ruan, Barbara Wold, Piero Carninci, Roderic Guigo, and Thomas R Gingeras. Landscape of transcription in human cells. *Nature*, 489(7414):101–108, sep 2012.

[112] Petr Danecek, Christoffer Nellaker, Rebecca E McIntyre, Jorge E Buendia-Buendia, Suzannah Bumpstead, Chris P Ponting, Jonathan Flint, Richard Durbin, Thomas M Keane, and David J Adams. High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. *Genome biology*, 13(4):26, apr 2012.

[113] Gregory R Grant, Michael H Farkas, Angel D Pizarro, Nicholas F Lahens, Jonathan Schug, Brian P Brunk, Christian J Stoeckert, John B Hogenesch, and Eric A Pierce. Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM). *Bioinformatics (Oxford, England)*, 27(18):2518–2528, sep 2011.

[114] Giacomo Baruzzo, Katharina E Hayer, Eun Ji Kim, Barbara Di Camillo, Garret A FitzGerald, and Gregory R Grant. Simulation-based comprehensive benchmarking of RNA-seq aligners. *Nature methods*, 14(2):135–139, feb 2017.

[115] Isaac D Raplee, Alexei V Evsikov, and Caralina Marín de Evsikova. Aligning the Aligners: Comparison of RNA Sequencing Data Alignment and Gene Expression Quantification Tools for Clinical Breast Cancer Research. *Journal of personalized medicine*, 9(2), apr 2019.

[116] Sung-Ho Goh, Matthew Josleyn, Y Terry Lee, Robert L Danner, Robert B Gherman, Maggie C Cam, and Jeffery L Miller. The human reticulocyte transcriptome. *Physiological genomics*, 30(2):172–178, jul 2007.

[117] Heesun Shin, Casey P Shannon, Nick Fishbane, Jian Ruan, Mi Zhou, Robert Balshaw, Janet E Wilson-McManus, Raymond T Ng, Bruce M McManus, and Scott J Tebbutt. Variation in RNA-Seq transcriptome profiles of peripheral whole blood from healthy individuals with and without globin depletion. *PloS one*, 9(3):e91041, 2014.

[118] ThermoFisher Scientific. GLOBINclear Rat mRNA Depletion Kit Protocol.

[119] AppliedBiosystems. GeneChip WT Pico Reagent Kit USER GUIDE.

[120] R N Van Gelder, M E von Zastrow, A Yool, W C Dement, J D Barchas, and J H Eberwine. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proceedings of the National Academy of Sciences of the United States of America*, 87(5):1663–1667, mar 1990.

[121] Benilton S Carvalho and Rafael A Irizarry. A framework for oligonucleotide microarray preprocessing. *Bioinformatics*, 26(19):2363–2367, 2010.

[122] Laurent Gautier, Leslie Cope, Benjamin M Bolstad, and Rafael A Irizarry. affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, 20(3):307–315, 2004.

[123] Patterned Flow Cells.

[124] Florent Hubé and Claire Francastel. Coding and Non-coding RNAs, the Frontier Has Never Been So Blurred. *Frontiers in genetics*, 9:140, 2018.

[125] Juliana Costa-Silva, Douglas Domingues, and Fabricio Martins Lopes. RNA-Seq differential expression analysis: An extended review and a software tool. *PLOS ONE*, 12(12):1–18, 2017.

[126] Zhijin Wu, Rafael A Irizarry, Robert Gentleman, Francisco Martinez-Murillo, and Forrest Spencer. A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. *Journal of the American Statistical Association*, 99(468):909–917, 2004.

[127] Yanli Zhang-James, Frank A Middleton, and Stephen V Faraone. Genetic architecture of Wistar-Kyoto rat and spontaneously hypertensive rat substrains from different sources. *Physiological Genomics*, 45(13):528–538, 2013.

[128] Dioze Guadagnini and Jose Antonio Rocha Gontijo. Altered renal sodium handling in spontaneously hypertensive rats (SHR) after hypertonic saline intracerebroventricular injection: role of renal nerves. *Life sciences*, 79(17):1666–1673, sep 2006.

[129] B M Sayers. Analysis of heart rate variability. *Ergonomics*, 16(1):17–32, jan 1973.

[130] G F Chess, R M Tam, and F R Calaresu. Influence of cardiac neural inputs on rhythmic variations of heart period in the cat. *The American journal of physiology*, 228(3):775–780, mar 1975.

[131] B W Hyndman, R I Kitney, and B M Sayers. Spontaneous rhythms in physiological control systems. *Nature*, 233(5318):339–341, oct 1971.

[132] S Akselrod, D Gordon, F A Ubel, D C Shannon, A C Berger, and R J Cohen. Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control. *Science (New York, N.Y.)*, 213(4504):220–222, jul 1981.

[133] Nina Japundzic-zigon. Physiological Mechanisms in Regulation of Blood Pressure Fast Frequency Variations. *Clinical and Experimental Hypertension*, 20(4):359–388, 1998.

[134] Nina Japundzić-Zigon. An update on blood pressure short-term variability. *TheScientificWorldJournal*, 2:320–323, feb 2002.

[135] Sanja Milutinovic, David Murphy, and Nina Japundzic-Zigon. The role of central vasopressin receptors in the modulation of autonomic cardiovascular controls: a spectral analysis study. *American journal of physiology. Regulatory, integrative and comparative physiology*, 291(6):R1579–91, dec 2006.

[136] Tatjana Tasic, Drago M Djordjevic, Silvio R De Luka, Alexander M Trbovich, and Nina Japundzic-Zigon. Static magnetic field reduces blood pressure short-term variability and enhances baro-receptor reflex sensitivity in spontaneously hypertensive rats. *International journal of radiation biology*, 93(5):527–534, may 2017.

[137] Olivera Šarenac, Maja Lozić, Srdja Drakulić, Dragana Bajić, Julian F Paton, David Murphy, and Nina Japundžić-Žigon. Autonomic mechanisms underpinning the stress response in borderline hypertensive rats. *Experimental physiology*, 96(6):574–589, jun 2011.

[138] Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *European heart journal*, 17(3):354–381, mar 1996.

[139] G Parati, J P Saul, M Di Rienzo, and G Mancia. Spectral analysis of blood pressure and heart rate variability in evaluating cardiovascular regulation. A critical appraisal. *Hypertension (Dallas, Tex. : 1979)*, 25(6):1276–1286, jun 1995.

[140] Mortara Andrea, La Rovere Maria Teresa, Pinna Gian Domenico, Prpa Alexander, Maestri Roberto, Febo Oreste, Pozzoli Massimo, Opasich Cristina, and Tavazzi Luigi. Arterial Baroreflex Modulation of Heart Rate in Chronic Heart Failure . *Circulation*, 96(10):3450–3458, nov 1997.

[141] Silke Ryan, Seamus Ward, Conor Heneghan, and Walter T McNicholas. Predictors of decreased spontaneous baroreflex sensitivity in obstructive sleep apnea syndrome. *Chest*, 131(4):1100–1107, apr 2007.

[142] M T La Rovere, G D Pinna, S H Hohnloser, F I Marcus, A Mortara, R Nohara, J T Jr Bigger, A J Camm, and P J Schwartz. Baroreflex sensitivity and heart rate variability in the identification of patients at risk for life-threatening arrhythmias: implications for clinical trials. *Circulation*, 103(16):2072–2077, apr 2001.

[143] Hanna Mussalo, Esko Vanninen, Risto Ikaheimo, Tomi Laitinen, Markku Laakso, Esko Lansimies, and Juha Hartikainen. Baroreflex sensitivity in essential and secondary hypertension. *Clinical autonomic research : official journal of the Clinical Autonomic Research Society*, 12(6):465–471, dec 2002.

[144] N Japundzic-Zigon. Effects of nonpeptide V1a and V2 antagonists on blood pressure fast oscillations in conscious rats. *Clinical and experimental hypertension (New York, N.Y. : 1993)*, 23(4):277–292, may 2001.

[145] Dragana Bajic, Tatjana Loncar-Turukalo, Sonja Stojicic, Olivera Sarenac, Tijana Bojic, David Murphy, Julian F R Paton, and Nina Japundzic-Zigon. Temporal analysis of the spontaneous baroreceptor reflex during mild emotional stress in the rat. *Stress (Amsterdam, Netherlands)*, 13(2):142–154, mar 2010.

[146] GraphPad Prism.

[147] Ralph Plehm, Marcos E Barbosa, and Michael Bader. Animal models for hypertension/blood pressure recording. *Methods in molecular medicine*, 129:115–126, 2006.

[148] Elena Wilde, Aisah A Aubdool, Pratish Thakore, Lineu Jr Baldissera, Khadija M Alawi, Julie Keeble, Manasi Nandi, and Susan D Brain. Tail-Cuff Technique and Its Influence on Central Blood Pressure in the Mouse. *Journal of the American Heart Association*, 6(6), jun 2017.

[149] R J Irvine, J White, and R Chan. The influence of restraint on blood pressure in the rat. *Journal of pharmacological and toxicological methods*, 38(3):157–162, nov 1997.

[150] *Spontaneously Hypertensive (SHR) Rats: Guidelines for Breeding, Care, and Use*. The National Academies Press, Washington, DC, 1976.

[151] Y Yamori, T Igawa, M Tagami, T Kanbe, Y Nara, M Kihara, and R Horie. Humoral trophic influence on cardiovascular structural changes in hypertension. *Hypertension (Dallas, Tex. : 1979)*, 6(6 Pt 2):III27–32, 1984.

[152] Krzysztof Narkiewicz and Guido Grassi. Impaired baroreflex sensitivity as a potential marker of cardiovascular risk in hypertension. *Journal of hypertension*, 26(7):1303–1304, jul 2008.

[153] Matheus M O Monteiro, Maria S Franca-Silva, Naiane F B Alves, Suenia K P Porpino, and Valdir A Braga. Quercetin improves baroreflex sensitivity in spontaneously hypertensive rats. *Molecules (Basel, Switzerland)*, 17(11):12997–13008, nov 2012.

[154] Mariane Bertagnolli, Cristina Campos, Paulo Cavalheiro Schenkel, Vera Lucia Longo de Oliveira, Katia De Angelis, Adriane Bello-Klein, Katya Rigatto, and Maria Claudia Irigoyen. Baroreflex sensitivity improvement is associated with decreased oxidative stress in trained spontaneously hypertensive rat. *Journal of hypertension*, 24(12):2437–2443, dec 2006.

[155] José R Cisternas, Vitor E Valenti, Thales B Alves, Celso Ferreira, Márcio Petenusso, João R Breda, Adilson C Pires, Nadir Tassi, and Luiz Carlos de Abreu. Cardiac baroreflex is already blunted in eight weeks old spontaneously hypertensive rats. *International archives of medicine*, 3:2, jan 2010.

[156] M C Andresen, S Kuraoka, and A M Brown. Baroreceptor function and changes in strain sensitivity in normotensive and spontaneously hypertensive rats. *Circulation research*, 47(6):821–828, dec 1980.

[157] Ryuji Nagai, Shinya Nagata, Fumiyo Fukuya, Jitsuo Higaki, Hiromi Rakugi, and Toshio Ogihara. Changes in autonomic activity and baroreflex sensitivity with the hypertension process and age in rats. *Clinical and Experimental Pharmacology and Physiology*, 30(5-6):419–425, 2003.

[158] Sean D Stocker and Martin S Muntzel. Recording sympathetic nerve activity chronically in rats: surgery techniques, assessment of nerve activity, and quantification. *American journal of physiology. Heart and circulatory physiology*, 305(10):H1407–16, nov 2013.

[159] Christopher R Pryce and Joram Feldon. Long-term neurobehavioural impact of the postnatal environment in rats: manipulations, effects and mediating mechanisms. *Neuroscience & Biobehavioral Reviews*, 27(1):57–71, 2003.

[160] Seymour Levine, David M Huchton, Sandra G Wiener, and Patricia Rosenfeld. Time course of the effect of maternal deprivation on the hypothalamic-pituitary-adrenal axis in the infant rat. *Developmental Psychobiology*, 24(8):547–558, 1991.

[161] G J Mangos, S W Turner, T B Fraser, and J A Whitworth. The role of corticosterone in corticotrophin (ACTH)-induced hypertension in the rat. *Journal of hypertension*, 18(12):1849–1855, dec 2000.

[162] Jonathan H Williams, Gordon K Wilcock, Jeffrey Seeburger, Aimee Dallob, Omar Laterza, William Potter, and A David Smith. Non-linear relationships of cerebrospinal fluid biomarker levels with cognitive function: an observational study. *Alzheimer's Research & Therapy*, 3(1):5, 2011.

[163] Mi Jeong Kwon, Ensel Oh, Seungmook Lee, Mi Ra Roh, Si Eun Kim, Yangsoon Lee, Yoon-La Choi, Yong-Ho In, Taesung Park, Sang Seok Koh, and Young Kee Shin. Identification of Novel Reference Genes Using Multiplatform Expression Data and Their Validation for Quantitative Gene Expression Analysis. *PLOS ONE*, 4(7):1–15, 2009.

[164] Hendrik J M de Jonge, Rudolf S N Fehrmann, Eveline S J M de Bont, Robert M W Hofstra, Frans Gerbens, Willem A Kamps, Elisabeth G E de Vries, Ate G J van der Zee, Gerard J te Meerman, and Arja ter Elst. Evidence based selection of housekeeping genes. *PloS one*, 2(9):e898, sep 2007.

[165] T W Kurtz, M Montano, L Chan, and P Kabra. Molecular evidence of genetic heterogeneity in Wistar-Kyoto rats: implications for research with the spontaneously hypertensive rat. *Hypertension (Dallas, Tex. : 1979)*, 13(2):188–192, feb 1989.

[166] W J Louis and L G Howes. Genealogy of the spontaneously hypertensive rat and Wistar-Kyoto rat strains: implications for studies of inherited hypertension. *Journal of cardiovascular pharmacology*, 16 Suppl 7:S1–5, 1990.

[167] W P Pare and J Kluczynski. Differences in the stress response of Wistar-Kyoto (WKY) rats from different vendors. *Physiology & behavior*, 62(3):643–648, sep 1997.

[168] E St Lezin, L Simonet, M Pravenec, and T W Kurtz. Hypertensive strains and normotensive 'control' strains. How closely are they related? *Hypertension (Dallas, Tex. : 1979)*, 19(5):419–424, may 1992.

[169] Bonnie Tocher Clause. The Wistar Institute Archives: Rats (Not Mice) and History. *Mendel Newsletter*, 1998.

[170] Gerard D'Angelo, Jennifer S Pollock, and David M Pollock. Endogenous endothelin attenuates the pressor response to acute environmental stress via the ETA receptor. *American journal of physiology. Heart and circulatory physiology*, 288(4):H1829–35, apr 2005.

[171] B T Alexander, K L Cockrell, A N Rinewalt, J N Herrington, and J P Granger. Enhanced renal expression of preproendothelin mRNA during chronic angiotensin II hypertension. *American journal of physiology. Regulatory, integrative and comparative physiology*, 280(5):R1388–92, may 2001.

[172] Mohamed A Bayorh, Agaba A Ganafa, Robin R Socci, Danita Eatman, Natalia Silvestrov, and Imad K Abukhalaf. Effect of losartan on oxidative stress-induced hypertension in Sprague-Dawley rats. *American journal of hypertension*, 16(5 Pt 1):387–392, may 2003.

[173] Debora S A Colombari, Eduardo Colombari, Andre H Freiria-Oliveira, Vagner R Antunes, Song T Yao, Charles Hindmarch, Alastair V Ferguson, Mark Fry, David Murphy, and Julian F R Paton. Switching control of sympathetic activity from forebrain to hindbrain in chronic dehydration. *The Journal of physiology*, 589(Pt 18):4457–4471, sep 2011.

[174] B J Sanders and J E Lawler. The borderline hypertensive rat (BHR) as a model for environmentally-induced hypertension: a review and update. *Neuroscience and biobehavioral reviews*, 16(2):207–217, 1992.

[175] J E Lawler, B J Sanders, R H Cox, and E F O'Connor. Baroreflex function in chronically stressed borderline hypertensive rats. *Physiology & behavior*, 49(3):539–542, mar 1991.

[176] M Burnier, J Biollaz, D B Brunner, and H R Brunner. Blood pressure maintenance in awake dehydrated rats: renin, vasopressin, and sympathetic activity. *American Journal of Physiology-Heart and Circulatory Physiology*, 245(2):H203–H209, 1983.

[177] R L Woods and C I Johnston. Contribution of vasopressin to the maintenance of blood pressure during dehydration. *The American journal of physiology*, 245(5 Pt 1):F615–21, nov 1983.

[178] S M Gardiner and T Bennett. Interactions between neural mechanisms, the renin-angiotensin system and vasopressin in the maintenance of blood pressure during water deprivation: studies in Long Evans and Brattleboro rats. *Clinical science (London, England : 1979)*, 68(6):647–657, jun 1985.

[179] Santosh S Atanur, Ana Garcia Diaz, Klio Maratou, Allison Sarkis, Maxime Rotival, Laurence Game, Michael R Tschannen, Pamela J Kaisaki, Georg W Otto, Man Chun John Ma, Thomas M Keane, Oliver Hummel, Kathrin Saar, Wei Chen, Victor Guryev, Kathirvel Gopalakrishnan, Michael R Garrett, Bina Joe, Lorena Citterio, Giuseppe Bianchi, Martin McBride, Anna Dominiczak, David J Adams, Tadao Serikawa, Paul Flicek, Edwin Cuppen, Norbert Hubner, Enrico Petretto, Dominique Gauguier, Anne Kwitek, Howard Jacob, and Timothy J Aitman. Genome sequencing reveals loci under artificial selection that underlie disease phenotypes in the laboratory rat. *Cell*, 154(3):691–703, aug 2013.

[180] Volker Fensterl and Ganes C Sen. Interferon-induced Ifit proteins: their role in viral pathogenesis. *Journal of virology*, 89(5):2462–2468, mar 2015.

[181] V R Antunes, S T Yao, A E Pickering, D Murphy, and J F R Paton. A spinal vasopressinergic mechanism mediates hyperosmolality-induced sympathoexcitation. *The Journal of physiology*, 576(Pt 2):569–583, oct 2006.

[182] Jing Qiu, Song Yao, Charles Hindmarch, Vagner Antunes, Julian Paton, and David Murphy. Transcription factor expression in the hypothalamo-neurohypophyseal system of the dehydrated rat: upregulation of gonadotrophin inducible transcription factor 1 mRNA is mediated by cAMP-dependent protein kinase A. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(9):2196–2203, feb 2007.

[183] Georgina G J Hazell, Charles C Hindmarch, George R Pope, James A Roper, Stafford L Lightman, David Murphy, Anne-Marie O'Carroll, and Stephen J Lolait. G protein-coupled receptors in the hypothalamic paraventricular and supraoptic nuclei–serpentine gateways to neuroendocrine homeostasis. *Frontiers in neuroendocrinology*, 33(1):45–66, jan 2012.

[184] Andreas Pichlmair, Caroline Lassnig, Carol-Ann Eberle, Maria W Gorna, Christoph L Baumann, Thomas R Burkard, Tilmann Burckstummer, Adrijana Stefanovic, Sigurd Krieger, Keiryn L Bennett, Thomas Rulicke, Friedemann Weber, Jacques Colinge, Mathias Muller, and Giulio Superti-Furga. IFIT1 is an antiviral protein that recognizes 5'-triphosphate RNA. *Nature immunology*, 12(7):624–630, jun 2011.

[185] Michael S Diamond and Michael Farzan. The broad-spectrum antiviral functions of IFIT and IFITM proteins. *Nature reviews. Immunology*, 13(1):46–57, jan 2013.

[186] Xiang Zhou, Jennifer J Michal, Lifan Zhang, Bo Ding, Joan K Lunney, Bang Liu, and Zhihua Jiang. Interferon induced IFIT family genes in host antiviral defense. *International journal of biological sciences*, 9(2):200–208, 2013.

[187] J Chebath, G Merlin, R Metz, P Benech, and M Revel. Interferon-induced 56,000 Mr protein and its mRNA in human cells: molecular cloning and partial sequence of the cDNA. *Nucleic acids research*, 11(5):1213–1226, mar 1983.

[188] J Kusari and G C Sen. Transcriptional analyses of interferon-inducible mRNAs. *Molecular and cellular biology*, 7(1):528–531, jan 1987.

[189] Ying Liu, Yi-Bing Zhang, Ting-Kai Liu, and Jian-Fang Gui. Lineage-Specific Expansion of IFIT Gene Family: An Insight into Coevolution with IFN Gene Family. *PLOS ONE*, 8(6):1–14, 2013.

[190] Koichiro Tamura, Joel Dudley, Masatoshi Nei, and Sudhir Kumar. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular biology and evolution*, 24(8):1596–1599, aug 2007.

[191] Volker Fensterl and Ganes C Sen. The ISG56/IFIT1 gene family. *Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research*, 31(1):71–78, jan 2011.

[192] Saumendra N Sarkar and Ganes C Sen. Novel functions of proteins encoded by viral stress-inducible genes. *Pharmacology & therapeutics*, 103(3):245–259, sep 2004.

[193] S J Gobin, A Peijnenburg, V Keijsers, and P J van den Elsen. Site alpha is crucial for two routes of IFN gamma-induced MHC class I transactivation: the ISRE-mediated route and a novel pathway involving CIITA. *Immunity*, 6(5):601–611, may 1997.

[194] S D Der, A Zhou, B R Williams, and R H Silverman. Identification of genes differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays. *Proceedings of the National Academy of Sciences of the United States of America*, 95(26):15623–15628, dec 1998.

[195] A Kohli, X Zhang, J Yang, R S Russell, R P Donnelly, F Sheikh, A Sherman, H Young, T Imamichi, R A Lempicki, H Masur, and S Kottilil. Distinct and overlapping genomic profiles and antiviral effects of Interferon-λ and -α on HCV-infected and noninfected hepatoma cells. *Journal of viral hepatitis*, 19(12):843–853, dec 2012.

[196] Luca D D'Andrea and Lynne Regan. TPR proteins: the versatile helix. *Trends in biochemical sciences*, 28(12):655–662, dec 2003.

[197] William C Merrick. Cap-dependent and cap-independent translation in eukaryotic systems. *Gene*, 332:1–11, may 2004.

[198] Daniel J Hui, C Ramana Bhasker, William C Merrick, and Ganes C Sen. Viral stress-inducible protein p56 inhibits translation by blocking the interaction of eIF3 with the ternary complex eIF2.GTP.Met-tRNAi. *The Journal of biological chemistry*, 278(41):39477–39482, oct 2003.

[199] Daniel J Hui, Fulvia Terenzi, William C Merrick, and Ganes C Sen. Mouse p56 blocks a distinct function of eukaryotic initiation factor 3 in translation initiation. *The Journal of biological chemistry*, 280(5):3433–3440, feb 2005.

[200] Fulvia Terenzi, Daniel J Hui, William C Merrick, and Ganes C Sen. Distinct induction patterns and functions of two closely related interferon-inducible human genes, ISG54 and ISG56. *The Journal of biological chemistry*, 281(45):34064–34071, nov 2006.

[201] J Guo, D J Hui, W C Merrick, and G C Sen. A new pathway of translational regulation mediated by eukaryotic initiation factor 3. *The EMBO journal*, 19(24):6891–6899, dec 2000.

[202] Ivan Topisirovic, Yuri V Svitkin, Nahum Sonenberg, and Aaron J Shatkin. Cap and cap-binding proteins in the control of gene expression. *Wiley interdisciplinary reviews. RNA*, 2(2):277–298, 2011.

[203] Fatima Gebauer and Matthias W Hentze. Molecular mechanisms of translational control. *Nature reviews. Molecular cell biology*, 5(10):827–835, oct 2004.

[204] Jing Shao, Masaomi Nangaku, Toshio Miyata, Reiko Inagi, Koei Yamada, Kiyoshi Kurokawa, and Toshiro Fujita. Imbalance of T-cell subsets in angiotensin II-infused hypertensive rats with kidney injury. *Hypertension (Dallas, Tex. : 1979)*, 42(1):31–38, jul 2003.

[205] Ashton Barnett-Vanes, Anna Sharrock, Mark A Birrell, and Sara Rankin. A Single 9-Colour Flow Cytometric Method to Characterise Major Leukocyte Populations in the Rat: Validation in a Model of LPS-Induced Pulmonary Inflammation. *PloS one*, 11(1):e0142520, 2016.

[206] Invitrogen. Lipofectamine 3000 Reagent Protocol.

[207] Zhenlin Yang, Huanhuan Liang, Qian Zhou, Ying Li, Haiwei Chen, Wen Ye, Danying Chen, Joy Fleming, Hongbing Shu, and Yingfang Liu. Crystal structure of ISG54 reveals a novel RNA binding structure and potential functional mechanisms. *Cell research*, 22(9):1328–1338, sep 2012.

[208] Abcam. Flow cytometry intracellular staining protocol.

[209] Jason E McDermott, Keri B Vartanian, Hugh Mitchell, Susan L Stevens, Antonio Sanfilippo, and Mary P Stenzel-Poore. Identification and validation of Ifit1 as an important innate immune bottleneck. *PloS one*, 7(6):e36465, 2012.

[210] Minghua Fan, Xiaobing Li, Xiaolin Gao, Lihua Dong, Gang Xin, Liqun Chen, Jianqing Qiu, and Yongping Xu. LPS Induces Preeclampsia-Like Phenotype in Rats and HTR8/SVneo Cells Dysfunction Through TLR4/p38 MAPK Pathway. *Frontiers in physiology*, 10:1030, 2019.

[211] Fernanda Brognara, Jaci Airton Castania, Daniel Penteado Martins Dias, Alexandre Kanashiro, and Helio Cesar Salgado. Time Course of Hemodynamic Responses to Different Doses of Lipopolysaccharide in Unanesthetized Male Rats. *Frontiers in physiology*, 10:771, 2019.

[212] Wilbur Y W Lew, Evelyn Bayna, Erminia Dalle Molle, Nancy D Dalton, N Chin Lai, Valmik Bhargava, Vincent Mendiola, Paul Clopton, and Tong Tang. Recurrent exposure to subclinical lipopolysaccharide increases mortality and induces cardiac fibrosis in mice. *PloS one*, 8(4):e61057, 2013.

[213] Peng Shi, Carlos Diez-Freire, Joo Yun Jun, Yanfei Qi, Michael J Katovich, Qiuhong Li, Srinivas Sriramula, Joseph Francis, Colin Sumners, and Mohan K Raizada. Brain microglial cytokines in neurogenic hypertension. *Hypertension (Dallas, Tex. : 1979)*, 56(2):297–303, aug 2010.

[214] A H de Haas, H R J van Weering, E K de Jong, H W G M Boddeke, and K P H Biber. Neuronal chemokines: versatile messengers in central nervous system cell interaction. *Molecular neurobiology*, 36(2):137–151, oct 2007.

[215] NICHOLAS J GAY and FIONNA J KEITH. Drosophila Toll and IL-1 receptor. *Nature*, 351:355, may 1991.

[216] Luke A J O'Neill, Douglas Golenbock, and Andrew G Bowie. The history of Toll-like receptors - redefining innate immunity., jun 2013.

[217] Jared C Roach, Gustavo Glusman, Lee Rowen, Amardeep Kaur, Maureen K Purcell, Kelly D Smith, Leroy E Hood, and Alan Aderem. The evolution of vertebrate Toll-like receptors. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27):9577 LP – 9582, jul 2005.

[218] Rahul B Dange, Deepmala Agarwal, Ryoichi Teruyama, and Joseph Francis. Toll-like receptor 4 inhibition within the paraventricular nucleus attenuates blood pressure and inflammatory response in a genetic model of hypertension. *Journal of Neuroinflammation*, 12(1):31, 2015.

[219] Robert F Schwabe, Ekihiro Seki, and David A Brenner. Toll-Like Receptor Signaling in the Liver. *Gastroenterology*, 130(6):1886–1900, 2006.

[220] Ekihiro Seki and David A Brenner. Toll-like receptors and adaptor molecules in liver disease: Update. *Hepatology*, 48(1):322–335, 2008.

[221] Dmitry A Chistiakov, Kenneth W Ashwell, Alexander N Orekhov, and Yuri V Bobryshev. Innervation of the arterial wall and its modification in atherosclerosis. *Autonomic neuroscience : basic & clinical*, 193:7–11, dec 2015.

[222] Jeffrey P Cardinale, Srinivas Sriramula, Nithya Mariappan, Deepmala Agarwal, and Joseph Francis. Angiotensin II-induced hypertension is modulated by nuclear factor-kappaBin the paraventricular nucleus. *Hypertension (Dallas, Tex. : 1979)*, 59(1):113–121, jan 2012.

[223] Yun-Ching Fu, Ching-Shiang Chi, Sui-Chu Yin, Betau Hwang, Yung-Tsung Chiu, and Shih-Lan Hsu. Norepinephrine induces apoptosis in neonatal rat cardiomyocytes through a reactive oxygen species-TNF alpha-caspase signaling pathway. *Cardiovascular research*, 62(3):558–567, jun 2004.

[224] Manveen K Gupta, T V Neelakantan, Mishra Sanghamitra, Rakesh K Tyagi, Amit Dinda, Subir Maulik, Chinmay K Mukhopadhyay, and Shyamal K Goswami. An assessment of the role of reactive oxygen species and redox signaling in norepinephrine-induced apoptosis and hypertrophy of H9c2 cardiac myoblasts. *Antioxidants & redox signaling*, 8(5-6):1081–1093, 2006.

[225] Mo-Lin Wang, Xiao-Jing Yu, Xiao-Guang Li, De-Zhi Pang, Qing Su, Roland Osei Saahene, Hong-Bao Li, Xu-Ying Mao, Kai-Li Liu, Li-Yan Fu, Ying Li, Guo-Qing Zhu, and Yu-Ming Kang. Blockade of TLR4 Within the Paraventricular Nucleus Attenuates Blood Pressure by Regulating ROS and Inflammatory Cytokines in Prehypertensive Rats. *American Journal of Hypertension*, 31(9):1013–1023, 2018.

[226] Qing Su, Jin-Jun Liu, Wei Cui, Xiao-Lian Shi, Jing Guo, Hong-Bao Li, Chan-Juan Huo, Yu-Wang Miao, Meng Zhang, Qing Yang, and Yu-Ming Kang. Alpha lipoic acid supplementation attenuates reactive oxygen species in hypothalamic paraventricular

nucleus and sympathoexcitation in high salt-induced hypertension. *Toxicology letters*, 241:152–158, jan 2016.

[227] Sveta Kabanova, Petra Kleinbongard, Jens Volkmer, Birgit Andree, Malte Kelm, and Thomas W Jax. Gene expression analysis of human red blood cells. *International journal of medical sciences*, 6(4):156–159, 2009.

[228] G W Schmid-Schonbein, D Seiffge, F A DeLano, K Shen, and B W Zweifach. Leukocyte counts and activation in spontaneously hypertensive and normotensive rats. *Hypertension (Dallas, Tex. : 1979)*, 17(3):323–330, mar 1991.

[229] Eric Vivier, Serge A van de Pavert, Max D Cooper, and Gabrielle T Belz. The evolution of innate lymphoid cells. *Nature immunology*, 17(7):790–794, jun 2016.

[230] Eric Vivier, Elena Tomasello, Myriam Baratin, Thierry Walzer, and Sophie Ugolini. Functions of natural killer cells. *Nature immunology*, 9(5):503–510, may 2008.

[231] Sabine Kossmann, Melanie Schwenk, Michael Hausding, Susanne H Karbach, Maria I Schmidgen, Moritz Brandt, Maike Knorr, Hanhan Hu, Swenja Kroller-Schon, Tanja Schonfelder, Stephan Grabbe, Matthias Oelze, Andreas Daiber, Thomas Munzel, Christian Becker, and Philip Wenzel. Angiotensin II-induced vascular dysfunction depends on interferon-gamma-driven immune cell recruitment and mutual activation of monocytes and NK-cells. *Arteriosclerosis, thrombosis, and vascular biology*, 33(6):1313–1319, jun 2013.

[232] Florent Ginhoux and Steffen Jung. Monocytes and macrophages: developmental pathways and tissue homeostasis. *Nature reviews. Immunology*, 14(6):392–404, jun 2014.

[233] Bernardo Rodriguez-Iturbe, Hector Pons, and Richard J Johnson. Role of the Immune System in Hypertension. *Physiological reviews*, 97(3):1127–1164, jul 2017.

[234] Dalit Strauss-Ayali, Sean M Conrad, and David M Mosser. Monocyte subpopulations and their differentiation patterns during infection. *Journal of leukocyte biology*, 82(2):244–252, aug 2007.

[235] Cord Sunderkotter, Tatjana Nikolic, Marilyn J Dillon, Nico Van Rooijen, Martin Stehling, Douglas A Drevets, and Pieter J M Leenen. Subpopulations of mouse blood monocytes differ in maturation stage and inflammatory response. *Journal of immunology (Baltimore, Md. : 1950)*, 172(7):4410–4417, apr 2004.

[236] Roxana Loperena, Justin P Van Beusecum, Hana A Itani, Noah Engel, Fanny Laroumanie, Liang Xiao, Fernando Elijovich, Cheryl L Laffer, Juan S Gnecco, Jonathan Noonan, Pasquale Maffia, Barbara Jasiewicz-Honkisz, Marta Czesnikiewicz-Guzik, Tomasz Mikolajczyk, Tomasz Sliwa, Sergey Dikalov, Cornelia M Weyand, Tomasz J Guzik, and David G Harrison. Hypertension and increased endothelial mechanical stretch promote monocyte differentiation and activation: roles of STAT3, interleukin 6 and hydrogen peroxide. *Cardiovascular Research*, 114(11):1547–1563, 2018.

[237] Kai-Uwe Belge, Farshid Dayyani, Alexia Horelt, Maciej Siedlar, Marion Frankenberger, Bernhard Frankenberger, Terje Espevik, and Loms Ziegler-Heitbrock. The proinflammatory CD14+CD16+DR++ monocytes are a major source of TNF. *Journal of immunology (Baltimore, Md. : 1950)*, 168(7):3536–3542, apr 2002.

[238] Leo M Carlin, Efstathios G Stamatiades, Cedric Auffray, Richard N Hanna, Leanne Glover, Gema Vizcay-Barrena, Catherine C Hedrick, H Terence Cook, Sandra Diebold, and Frederic Geissmann. Nr4a1-dependent Ly6C(low) monocytes monitor endothelial cells and orchestrate their disposal. *Cell*, 153(2):362–375, apr 2013.

[239] Amado Quintar, Sara McArdle, Dennis Wolf, Alex Marki, Erik Ehinger, Melanie Vassallo, Jacqueline Miller, Zbigniew Mikulski, Klaus Ley, and Konrad Buscher. Endothelial Protective Monocyte Patrolling in Large Arteries Intensified by Western Diet and Atherosclerosis. *Circulation research*, 120(11):1789–1799, may 2017.

[240] Prakash Babu Narasimhan, Paola Marcovecchio, Anouk A J Hamers, and Catherine C Hedrick. Nonclassical Monocytes in Health and Disease. *Annual review of immunology*, 37:439–456, apr 2019.

[241] Cell Analysis Facility. Cell Sorting for RNA Isolation.

[242] M T Barrett, J Glogovac, L J Prevo, B J Reid, P Porter, and P S Rabinovitch. High-quality RNA and DNA from flow cytometrically sorted human epithelial cells and tissues. *BioTechniques*, 32(4):888–890,892,894,896, apr 2002.

[243] Christine Vogel and Edward M Marcotte. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature reviews. Genetics*, 13(4):227–232, mar 2012.

[244] Helga Thorvaldsdottir, James T Robinson, and Jill P Mesirov. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics*, 14(2):178–192, mar 2013.

[245] Travis C Glenn. Field guide to next-generation DNA sequencers. *Molecular ecology resources*, 11(5):759–769, sep 2011.

[246] Jay Shendure and Hanlee Ji. Next-generation DNA sequencing. *Nature biotechnology*, 26(10):1135–1145, oct 2008.

[247] Zhenhua Hu, Shengmin Mei, Jie Xiang, Jie Zhou, Zhiwei Li, Lin Zhou, Sheng Yan, Weilin Wang, and Shusen Zheng. Survival rates after liver transplantation using hypertensive donor grafts: an analysis of the Scientific Registry of Transplant Recipients database. *Journal of hepato-biliary-pancreatic sciences*, 24(8):441–448, aug 2017.